Genomic approaches to unravel the pathogenesis of Chagas disease

Doctoral Thesis

Desiré Casares Marfil, 2021

Programa de Doctorado en Biomedicina

Tesis Doctoral

Genomic approaches to unravel the pathogenesis of Chagas disease

Memoria presentada por la graduada en Biología Desiré Casares Marfil para optar al grado de Doctora por la Universidad de Granada.

Programa de Doctorado en Biomedicina.

Directores:

Javier Martín Ibáñez, Profesor de Investigación del CSIC.

Marialbert Acosta Herrera, Investigadora Postdoctoral del Instituto Parasitología y Biomedicina López-Neyra.

Instituto de Parasitología y Biomedicina López-Neyra, CSIC.

Granada, noviembre de 2021

Editor: Universidad de Granada. Tesis Doctorales
Autor: Desiré Casares Marfil
ISBN: 978-84-1117-215-8
URI: http://hdl.handle.net/10481/72459

A José,

A mis padres y mi hermano,

A mis amigos y compañeros.

INDEX

ABBREVIATIONS

1KGP: The 1000 Genomes Project

AFR: African population

AMR: Native American population

CCR5: Chemokine receptor 5

CI: Confidence interval

CNV: Copy number variation

DC: Dendritic cell

DMP: Differentially methylated position

DMR: Differentially methylated region

DTG: Disease tolerance genes

DTUs : Discrete typing units

ELISA: Enzyme-linked immunosorbent assay

eQTL: Expression-quantitative trait *loci*

EUR: European population

FDR: False discovery rate

GRS: Genomic risk score

GWAS: Genome-wide association studies

HLA: Human leukocyte antigen

HP: Haptoglobin

HWE: Hardy-Weinberg equilibrium

IBD: Identity by descent

IFN: interferon

IL: Interleukins

LD: Linkage disequilibrium

MAF: Minor allele frequency

MHC: Major histocompatibility complex

MIF: Migration inhibitory factor

mQTL: Methylation quantitative trait *loci*

NFAT: Transcriptional nuclear factor of activated T cells

NK: Natural killer cells

NO : Nitric oxide

OR: Odds ratios

PCA: Principal component analysis

PCR: Polymerase chain reaction

PEL: Peruvian from Lima population

PheWAS: Phenome-Wide Association Studies

PI_HAT: IBD proportion

PRG: Pathogen resistance genes

QC: Quality control

RNA-seq: RNA-sequencing

SNPs: Single nucleotide polymorphism

TF: Transcription factor

TIRAP: Toll-interleukin-1 receptor domain containing adapter protein

TLRs: Toll-like receptors

TRIM: Tripartite Motif family

TSS: Transcription start site

WHO : World Health Organization

GENES

APOL1: Apolipoprotein L1

ATP2B4: ATPase Plasma Membrane Ca2+ Transporting 4

BAFT2: Basic Leucine Zipper ATF-Like Transcription Factor 2

CCDC88B: coiled-coil domain containing 88B

CCL2: C-C Motif Chemokine Ligand 2

CCL3: C-C Motif Chemokine Ligand 3

CCL4: C-C Motif Chemokine Ligand 4

CCL5: C-C Motif Chemokine Ligand 5

CCR2: C-C Motif Chemokine Receptor 2

CCR5: C-C Motif Chemokine Receptor 5

CD247: Cluster of Differentiation 247

CD40L: CD40 Ligand

CR1: Complement C3b/C4b Receptor 1

CXCL10: C-X-C Motif Chemokine Ligand 10

CXCL9: C-X-C Motif Chemokine Ligand 9

DUSP22: Dual Specificity Phosphatase 22

EBF2: EBF Transcription Factor 2

FERMT3: FERM Domain Containing Kindlin 3

FL1: Follicular Lymphoma, Susceptibility To, 1

HLA: Human leukocyte antigen

IFN-γ: Interferon gamma

IL-10: Interleukin 10

IL-12: Interleukin 12

IL-15: Interleukin 15

IL-17: Interleukin 17

IL-17A: Interleukin 17A

IL-18: Interleukin 18

IL-2: Interleukin 2

IL-35: Interleukin 35

IL-6: Interleukin 6

IL-7: Interleukin 7

KIR3DS1: Killer Cell Immunoglobulin Like Receptor, Three Ig Domains And Short Cytoplasmic Tail 1

KLF4: Kruppel Like Factor 4

KLF4: Kruppel-like factor 4

LAMP3: Lysosomal Associated Membrane Protein 3

LTA: Lymphotoxin Alpha

MBL2: Mannose-binding lectin 2

NAALADL1: N-Acetylated Alpha-Linked Acidic Dipeptidase Like 1

NLRP3: NLR pyrin domain-containing 3

PLA2G16: Alias for *PLAAT3* gene

PLAAT3: Phospholipase A and acyltransferase 3

PLCB3: Phospholipase C Beta 3

POLA2: DNA polymerase alpha 2, accessory subunit

RPP21: Ribonuclease P/MRP Subunit P21

SAC3D1: SAC3 domain-containing protein 1

SHD1: Alias for *SAC3D1* gene

SIK1: Salt Inducible Kinase 1

SLC11A1: Solute Carrier Family 11 Member 1

SNX15: Sorting Nexin 15

STAT5: Signal transducer and activator of transcription 5

STX7: Syntaxin 7

TNF-α: Tumor necrosis factor

TRIM10: Tripartite Motif Containing 10

TRIM26: Tripartite Motif Containing 26

TRIM31: Tripartite Motif Containing 31

TRIM39: Tripartite Motif Containing 39

TRIM40: Tripartite Motif Containing 40

VTRNA2-1: Vault RNA 2-1

RESUMEN

La enfermedad de Chagas es una enfermedad infecciosa causada por el parásito *Trypanosoma cruzi*. Esta enfermedad es endémica de Latinoamerica donde el principal vector de transmisión son insectos hematófagos. Se estima que afecta a unas 6-7 millones de personas en todo el mundo ya que, debido a procesos como la migración y globalización, se ha extendido a zonas no endémicas. La enfermedad cursa con una fase aguda seguida de una fase indeterminada en la que la mayoría de individuos permanecen asintomáticos de por vida. Sin embargo, algunos pacientes pueden desarrollar la forma crónica de la enfermedad años después de la infección, manifestándose con sintomatología digestiva y/o cardíaca, siendo esta última la más frecuente y grave conocida como cardiomiopatía chagásica crónica.

Dada las diferencias en el desarrollo de la infección en zonas endémicas así como la evolución diferencial a la forma crónica cardíaca, se ha sugerido la influencia de la variación genética del hospedador en la patogénesis de la enfermedad. La presente tesis doctoral se centra en el estudio del componente genético del hospedador mediante diferentes estrategias de análisis genético, para dilucidar su papel tanto en la infección como en el desarrollo de la cardiomiopatía chagásica crónica.

Para ello y con el objetivo de confirmar asociaciones previamente descritas, se realizaron estudios de genes candidatos para evaluar la relación de variantes genéticas localizadas en los genes *IL6*, *IL17A* e *IL18*, que codifican citoquinas implicadas en la respuesta inmune contra *Trypanosoma cruzi.* Dichas asociaciones se confirmaron para los genes *IL17A* e *IL18*, mientras que la relación de la variante del gen *IL6* fue descartada.

Sin embargo, la evaluación de variantes genéticas a lo largo de todo el genoma mejora la capacidad de identificación de variación asociada con la enfermedad de una manera no sesgada, para lo cual un estudio de asociación del genoma completo (GWAS, del inglés "genome-wide association studies") supuso la herramienta de elección. Así, se realizó un GWAS en individuos procedentes de Colombia, Bolivia y Argentina que, junto con la información de población brasileña previamente publicada, se meta-analizó tanto para la susceptibilidad a la infección como para el desarrollo de la cardiomiopatía chagásica crónica. Este análisis nos permitió identificar una asociación a nivel de significación genómica con la forma cardíaca cercana al gen *SAC3D1*, siendo la primera asociación genómica identificada en esta enfermedad.

Por otro lado, dada la mezcla de las poblaciones ancestrales Nativo Americana, Europea y Africana presente en las poblaciones latinas, se evaluó la relación de estos bloques de ascendencia con la susceptibilidad a la enfermedad de Chagas. Para esto se realizó un análisis de mapeo por mezcla en la cohorte colombiana. Se identificó una asociación de protección de la ascendencia Nativo Americana con la susceptibilidad a la infección en la región del complejo mayor de histocompatibilidad, en contraposición de la ascendencia Europea que resultó de riesgo, poniendo de manifiesto el componente evolutivo en la relación parásito-hospedador de estas poblaciones. En esta región genética se localizan numerosos genes implicados en la respuesta inmunológica, entre los que cabría destacar los genes *HLA*, que tienen una gran transcendencia en el reconocimiento antigénico, y algunos miembros de la familia de los *TRIM*, conocidos por su papel en la respuesta inmune frente a patógenos.

Finalmente, se llevó a cabo un análisis de rasgos cuantitativos de metilación (mQTLs, del inglés "methylation quantitative trait *loci*") en la región genómica identificada en el GWAS. Con esta estrategia se analizó

el efecto de la variación genética en los patrones de metilación de un subconjunto de pacientes procedentes de Bolivia. Entre los resultados más importantes destacamos aquellos mQTLs en genes que se han relacionado previamente con caracteres hematológicos y cardiovasculares. Además, identificamos mQTLs del gen *CCDC88B,* que tiene un papel relevante en el proceso inflamatorio. Este gen había sido previamente identificado como diferencialmente metilado y expresado en tejido cardiaco de pacientes con cardiomiopatía chagásica crónica, lo cual pone de manifiesto su importancia en la patogénesis de la enfermedad. Nuestros resultados confirmaron la influencia de los genes de esta región, destacando la funcionalidad de los mismos más allá de la variación en la secuencia del ADN. Adicionalmente, la identificación de genes previamente identificados en tejido cardiaco, pone de manifiesto la importancia del uso de tejidos más accesibles en la identificación de biomarcadores.

Los resultados presentados en esta tesis doctoral destacaron el papel de la genética del hospedador en el desarrollo de la enfermedad de Chagas. Esto, junto con los análisis de metilación y expresión global que se están realizando en estas muestras, representan un avance en el conocimiento de la patogénesis de esta enfermedad tropical desatendida.

ABSTRACT

Chagas disease is an infection caused by the parasite *Trypanosoma cruzi*. This disease is endemic form Latin America, where the main transmission vectors are hematophagous bugs. It is estimated that the infection affects 6-7 million people around the world, where it has been extended to non-endemic areas given migratory movements and globalization. Chagas disease has an acute phase followed by an indeterminate phase, and the majority of individuals remain asymptomatic for life. Nevertheless, some patients could develop the chronic form of the disease even years after the infection, which manifests with digestive and/or cardiac symptomatology. The latter the most frequent and severe form of the disease is known as chronic Chagas cardiomyopathy.

Given the inter-individual differences in the infection in endemic regions as well as the differential development of the chronic cardiac form, it has been suggested the influence of the host genetic component in the pathogenesis of the disease. The present doctoral thesis focused in the analysis of the host genetic component through different genetic analysis approaches, in order to elucidate its role in the infection and the progress to the chronic Chagas cardiomyopathy.

In order to confirm previously described associations, candidate gene studies were performed to assess genetic variants located in the *IL6*, *IL17A* and *IL18* genes, which encode cytokines with relevance in the immune response against *Trypanosoma cruzi*. The associations were confirmed for the *IL17A* and *IL18* variants, while it was discarded for the *IL6* gene variant.

The assessment of genetic variation throughout the entire genome improves the capacity of identifying disease associations in an unbiased fashion; therefore a genome-wide association study (GWAS) was the strategy of choice. For this, Colombian, Bolivian and Argentinian individuals' genomic data was meta-analyzed with previously published data from the Brazilian population. This analysis was carried out for the susceptibility to the infection and the development of chronic Chagas cardiomyopathy. Using this strategy, it was possible to identify an association at genomic significance level with the chronic cardiac form in the vicinity of the *SAC3D1* gene, being this the first genomic association identified for the disease.

On the other hand, given the admixture of Native American, European and African ancestries of the Latin American populations, the relation among ancestry genetic blocks and the susceptibility to Chagas disease was assessed. Thus, an admixture mapping analysis in the Colombian population was performed. This approach revealed a protective association of the Native American ancestry with the susceptibility to the infection located in the major histocompatibility complex region. In contrast, the European ancestry showed a risk association with the disease, highlighting the evolutionary component in the parasite-host relationship of these populations. The HLA genes and the TRIM gene family were highlighted in this locus, which are of great importance for antigenic recognition and their contribution in the immune response against pathogens.

Finally, a methylation quantitative trait *loci* (mQTLs) analysis was performed in the significant genomic region from the GWAS. This strategy allowed us to evaluate the effect of the genetic variation in the methylation patterns in a subset of individuals from Bolivia. Significant

mQTLs were identified, including the *CCDC88B*, which is an important player in the inflammatory process. Interestingly, a previous analysis in chronic Chagas cardiomyopathy patients revealed this gene as differentially methylated and expressed. Our results highlighted the functionality of the genes within this region beyond the variation in the DNA sequence. In addition, the exposure of genes previously reported in cardiac samples, highlighted the importance of using more accessible tissues in the identification of disease biomarkers.

The results presented in this doctoral thesis emphasized the role of the host genetics in the development of Chagas disease. This, together with the global methylation and expression analyses being carried out, represents an advance in the knowledge of the pathogenesis of this neglected tropical disease.

INTRODUCTION

1. Chagas disease: a neglected tropical disease

Chagas disease, also known as American trypanosomiasis, is an infection caused by the parasitic protozoan *Trypanosoma cruzi*, whose main transmission vector are bugs from the Reduviidae family (1). The disease is endemic in 21 Latin American countries and, according to the last World Health Organization (WHO) report, it is estimated that affects 6-7 million people around the world (https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis)). Although the infection was restricted to rural areas from Latin America at the beginning, the migratory movements have originated the globalization of the disease, extending it to urban areas and other regions such as USA or Europe (Figure 1) (2).

Figure 1. Regions with cases of Chagas disease in endemic and non-endemic areas (Modified from Sangenito et al, Trop Med Infect Dis. 2020).

The infection has been dated in the pre-Columbian period thanks to the discovery of *T. cruzi* DNA in mummies from Chile/Peru and Brazil as far back as 7,500 BC (3). However, it was not until 1909 that the disease, the ethological agent and the bug vector that transmit it were described for the first time by the Brazilian physician Carlos Ribeiro Justiniano Chagas (1879–1934) in a 2 years-old patient with presence of the parasite in blood (4). Thus, it is believed that the great variability of these insect vectors, as well as the animal reservoirs, have contributed to the persistence of the infection for almost 9,000 years (5).

Over time, other transmission ways of the parasite have diversified, being even more common than the Triatomine bugs in certain endemic regions (6). The transmission by blood transfusion started to be less common in endemic areas after its detection, however it might suppose a challenge in non-endemic regions where blood donations are not screen tested for the presence of the parasite (7). Oral transmission has also been described in Chagas disease, mainly through wild animal meat and the consumption of food contaminated with the feces of the insect vector (6, 8). Moreover, this transmission mechanism has been described to have higher mortality rate in comparison to the vector transmission (8).

After the entry of the parasite, Chagas disease starts with an acute phase, which can be followed by an indeterminate phase that in most cases uses to be asymptomatic (9). However, some Chagas patients develop chronic symptoms that compromise internal organs such as heart, esophagus and colon, comprising this symptomatology the chronic cardiac and digestive forms of the disease, respectively (https://www.cdc.gov/parasites/chagas/disease.html). In spite of the variety of symptoms that Chagas patients can manifest, a high number of them remain undiagnosed and, in combination with the high transmission

rate, it is estimated that 75 million people are in risk of infection (https://www.who.int/health-topics/chagas-disease#tab=tab_1). This produces patient's disability and long-term treatments in the case of the chronic forms, which make Chagas disease one of the most costly neglected tropical diseases (1). In addition, the high mobility emphasizes the economic burden of Chagas disease in non-endemic regions where it is estimated that a 10% of its total burden affects these regions (1, 9, 10).

2. *Trypanosoma cruzi* life cycle

The life cycle of *T. cruzi* takes place in the main transmission vector, the Triatomine bugs, and in the animal host (11). During this complex cycle, the parasite goes through four different stages: blood-form trypomastigotes, intracellular amastigotes, epimastigotes and metacyclic trypomastigotes (Figure 2) (11). The cycle starts when the insect vector ingests the blood-form trypomastigotes from an infected host through its bite. Once in the mammalian host, the parasite starts division cycles by binary fission in the vector's gut, being this the epimastigote parasitic form (12). Next, the parasite is differentiated into metacyclic trypomastigotes, which migrates to the hindgut of the vector around three weeks later. Normally, the transmission to the mammalian host occurs when the Triatomine vector bites and deposits its feces in the host's skin, which favors the entry of the parasite through the wound (12). Metacyclic trypomastigotes will penetrate into the host' cells to turn into the amastigotes stage, which uses binary fission to multiply and infect the tissue (11). At this point, the host will suffer the acute phase of the disease while the number of amastigotes increases and infect other tissues, turning into the trypomastigotes stage (9, 12). The last ones will reside in blood where they can be transmitted to another bug, closing the life cycle of the parasite (11).

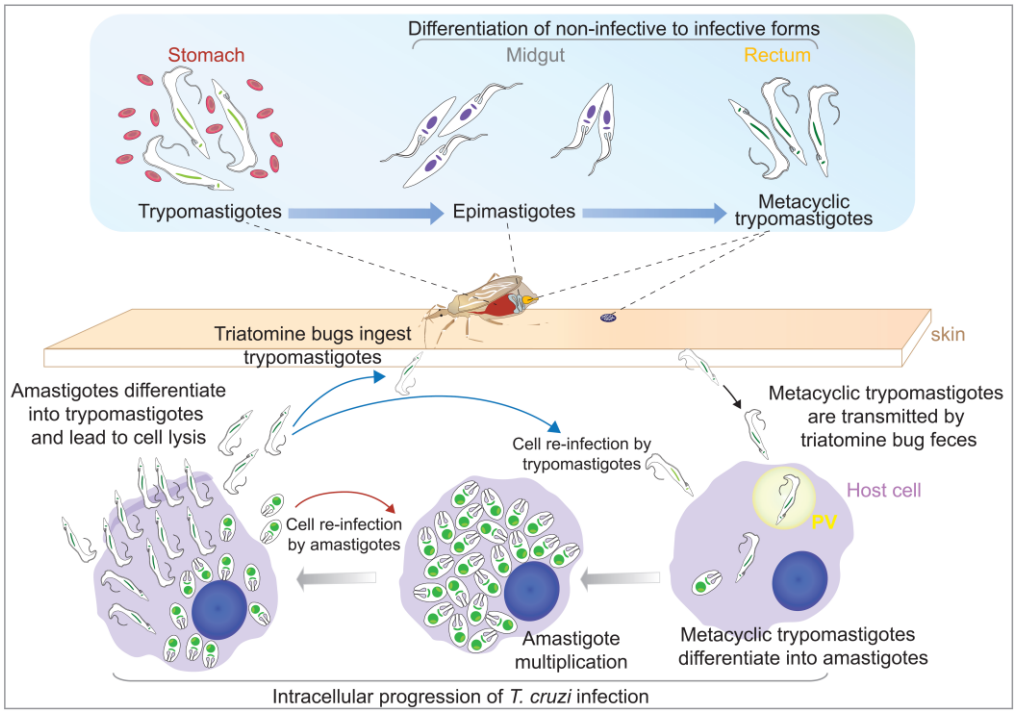

Figure 2. *Trypanosoma cruzi* double-host cycle (Modified from Moretti et al, Trends Parasitol. 2020).

Six phylogenetic lineages or discrete typing units (DTUs) have been described for *T. cruzi* named TcI to TcVI and classified according its genetics, geographical distribution, pathogenesis and response to treatment (9, 13). In addition, different DTUs have been associated with different symptomatology of the disease, for instance the TcII has been related with cardiac complications of the chronic phase while the TcI has been mostly identified in individuals with acute manifestations of the disease (14).

During this cycle, there are several metabolic processes that favor the parasite replication and might affect the pathogenesis of the disease (12).

One of these processes is the variation of ionic concentrations that affect to the different parasitic stages (12). The calcium (Ca^{2+}) and potassium (K^{+}) have been described to play an important role in the proliferation and the tissue invasion of the parasite in the human host through the invagination of the cellular membrane (11, 12, 15, 16). Thus, the regulation of these molecules by ionic channels allows to the parasite the adaptation to changing conditions in the vector and the mammalian host (17). In fact, parasite's calcium homeostasis has been proposed as a possible therapeutic target for possible treatments of the disease (18). Another molecule is the nitric oxide (NO), which is involve in the control of parasitic levels through toxic levels and the reduction of growth factors (19).

3. Pathogenesis of Chagas disease

During the acute phase, the first contact with the parasite triggers host immune response that might develop nonspecific clinical symptoms such as fever and inflammation (20). Nevertheless, some of these symptoms can help to the detection of the infection in endemic regions, for example when the parasite entry through the conjunctiva and causes eye inflammation, which is known as the Romaña sign (21). This phase uses to appear a few weeks after the infection with the parasite and last for 4-8 weeks (9). About 1-5% of patients may develop a severe acute phase characterized by myocarditis and/or meningoencephalitis, which can lead to death (4, 9). During the acute phase the screening of seropositive patients is very important for an early administration of treatment in order to palliate symptoms and prevent the chronic phase (22). Thus, detection methods are based on the identification of the parasite and change according to the stage of the disease (23). Therefore, during the acute phase, methods based on concentration or PCR (polymerase chain reaction) techniques predominate,

while in the chronic phase serological methods such as ELISA (Enzyme-linked immunosorbent assay) are used for screening (23). This supposes key in the control of the disease as most patients remain asymptomatic during the acute phase and for the rest of their lives, being this known as the indeterminate phase (9). Nevertheless, around 30-40% of infected patients develop a chronic phase with organ involvement that can appear even years after contact with the parasite (4). This chronic phase can produce cardiomyopathy, also known as chronic Chagas cardiomyopathy, or involve the digestive system and produce megaoesophagus and megacolon (4, 24).

The chronic Chagas cardiomyopathy is the most severe form of the disease, being the sudden death the main cause of death at this stage of the disease (4, 24-26). This chronic phase affects mainly the myocardium and the conduction system, although it can also affect the circulatory system causing other symptoms such as thrombus formation (27). According to the severity of symptoms, there are different classifications for the chronic cardiac form. These are based on the presence of electrocardiogram abnormalities and its severity (28). Some of these classifications are the Kuschnir and that according to the American Heart Association (28-30). In general, these classify patients into 4 groups based on the results of the echocardiogram and electrocardiogram, as well as physiological abnormalities in the tissue or organ (28, 31).

The chronic digestive form of Chagas disease is less frequent among chronic patients although it can also appear in conjunction with the chronic cardiac form (32). Symptoms of this condition range from dysphagia and motility problems of the digestive track to more severe forms such as megacolon and megaesophagus (33). The digestive affection is mainly reported in regions from Brazil being also related with specific parasitic DTUs, although it has become more common in non-endemic regions (32).

In this sense, several studies have related the geographical distribution of *T. cruzi* DTUs with the different clinical symptoms prevalence in different regions (34). Thus, the TcI and TcIV strains have been mainly detected among acute cases of Chagas disease while the TcII presented highly prevalence in different cardiac and digestive affections in patients from Brazil (13).

The variety of parasitic strains has also been shown to affect the efficacy of the main treatments currently used, benznidazole and nifurtimox, due to their different resistance (35). Benznidazole and nifurtimox are antiparasitic drugs with better results observed in young patients and during the early stages of the disease, since in the advanced chronic phase and patients of advanced age the side effects are increased (9, 35). For this reason, and especially for the acute phase of the disease, the development of new, less invasive and individualized treatments is necessary. In this sense, precision medicine allows the identification of new markers and therapeutic targets of disease progression useful for the development of new drugs (36).

3.1. Host immune response

The parasitic control and disease progression during the *T. cruzi* infection depend on the innate and adaptive immune responses (Figure 3), as well as of several external factors , such as genetic parasitic variation, the transmission way, neuroendocrine factors or the host's genetic background (37, 38). In the early infection the high parasitic levels requires a strong inflammatory response in order to control the parasite distribution to other tissues (9). Thus, in the acute phase mainly highlight the role of macrophages, natural killer cells (NK), lymphocytes and cytokines production (38).

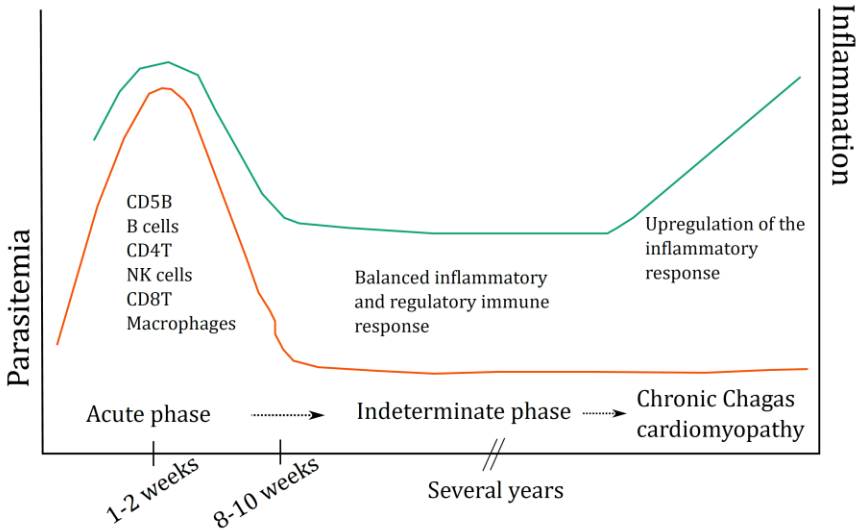

Figure 3. During Chagas disease stages there are different balances between the regulatory and inflammatory immune response as parasite levels in blood or tissues change (parasitemia). Thus, after the entry of the parasite, immune cells are produced in order to control the infection. This originated a balance in the immune system that keeps most individuals asymptomatic for life. However, in 30% of

patients there is an imbalance in the inflammatory state, characterized by the production of proinflammatory cytokines, which affects the cardiomyocytes causing the chronic Chagas cardiomyopathy (Modified from Machado et al, Semin Immunopathol. 2012).

In the innate immunity, the first line of defense is composed of monocytes/macrophages, neutrophils and dendritic cells due to their disposition in the epithelial region (39). In the acute phase, monocytes and neutrophils express a series of receptors specialized in antigen recognition, the most prominent of which are the Toll-like receptors (TLRs) (39). Several *T. cruzi* molecules act as ligand to these receptors, thus, it has been reported that *T. cruzi* DNA and RNA act as activators for the TLR9 and 7, respectively (40). In addition, during the innate immune response, the release of cytokines seems to be important both in the initial infection and in the progression of the disease (41). Cytokines such as interleukins 6 and 12 (IL-6 and IL-12), interferon gamma (IFN-γ) and the tumor necrosis factor (TNF-α), among others, are the first to be released and responsible for the activation of other cells of the inflammatory response (36, 42). One of these cells activated by cytokines are the NK cells that have a crucial role during intracellular infections by controlling parasitic levels during early stages of Chagas disease (39). Although these cells exert their main function by releasing IFN-γ, they can also fight infection by eliminating extracellular parasites in an IL-12 dependent mechanism, or by influencing the maturation of other cells such as dendritic cells (DCs) (39). DCs are also part of the first line of defense against the parasite playing an important role during the acute phase (43). In this sense, several studies propose these cells as a treatment therapy due to their ability to recognize and internalize antigens during infection, specifically tolerogenic dendritic cells, which have an inhibitory effect in the progression of chronic Chagas cardiomyopathy (43, 44).

During the adaptive immune response, B and T cells play a fundamental role, since antigen recognition and antibody production are crucial processes for the control of the parasite in the organism and to maintain the survival rate (45). The B cells are the main producers of antibodies against *T. cruzi*, providing a bridge between innate and adaptive immune responses (39, 46). Numerous studies have shown that the antibody production is crucial in the course of Chagas disease during the acute and chronic phases (45, 47). Nevertheless, the antibodies produced are not fully efficient eliminating the parasite (39). In this sense, a continued patients' exposure to the parasite, although in low levels, might produces an inflammatory response during the chronic phase that promotes cardiac damage (24). In addition to the production of antibodies, different types of B cells produce a variety of cytokines that stimulates the production of T cells. For example, regulatory B cells produce IL-10 and IL-35, which stimulate the formation of T cell clones as well as promote regulatory T cells (48). Thus, the T cells play an important role in the adaptive immune response against *T. cruzi*, as they are the main responsible for the antigen recognition from the antigen presenting cells during the acute phase (39). In the case of the chronic phase, these cells also produce different types of cytokines that helps to control parasitic levels and the damage level (49).

The immune response during the first stage of the disease is crucial to prevent parasite proliferation in tissues and the development of the chronic phase (4). Thus, it has been shown that chronic Chagas patients have higher levels of T cells with immunoregulatory activity (49). However, the expression of TNF-α and IFN-γ cytokines by these cells has been associated with the development of a more severe chronic cardiac form probably as a consequence of a continuous exposure to the parasite, which has been proposed as a promoter of cardiac damage in Chagas patients (24, 49). This is the premise of the autoimmunity in Chagas disease hypothesis, which is

proposed due to the presence of autoantibodies in the course of the disease (50). Different autoantibodies, such as targets of β1-adrenergic receptors, have been related with the chronic cardiac form of Chagas disease supporting the idea of anti-self-response that has been also related with parasitic persistence (51). Thus, the fact that better explain this hypothesis is the existence of a molecular mimicry between parasitic and host molecules (50, 52). Nevertheless, even though the existence of autoantibodies has been demonstrated in Chagas disease and patients with the chronic cardiac form, their implication in the development of the disease needs to be clarified (51, 52).

4. The study of the genetic component of diseases.

Genetic studies are based on the identification of association among a specific trait or disease with genetic variation. This variation in the DNA sequence goes from copy number variations (CNVs) or microsatellites to single nucleotide polymorphisms (SNPs), being the last one the most common markers under study (53).

The SNPs are variations in a single position of the DNA sequence with a minor allele frequency (MAF) of at least 1% in the population under study. It is estimated that there are around 38 million SNPs across the genome and, although the majority are in non-coding regions, they compound around 90% of the human genetic variation (54). Genetic analyses allow to evaluate the relation of this genetic variation with a trait or disease given us an idea of the genetic predisposition of a group of individuals to develop it (https://www.nature.com/scitable/definition/snp-295/). In this sense, allelic frequencies of one or more SNPs are compared among cases (individuals affected with the disease) and controls (unaffected individuals)

using different statistical methods. Thus, if they are statistically significant under different significance thresholds, they are considered in association with the disease (55).

The first significant event in genomics research was the sequencing of the human genome, driven by the Human Genome Project in 1990 (56). This announcement would accelerate biomedical research, allowing the determination of the genetic contribution to human traits (57). From this, candidate gene studies emerged as a rapid and cheap first approach by which the genetic relatedness of genes to a trait or disease is assessed by selecting SNPs located in their gene region (58). These types of studies are performed under a previous association hypothesis mainly based on the functional implication of the gene (59).

The determination of common patterns of the DNA sequence variation in different populations by the International HapMap Project would contribute to the genomic development with the determination of the linkage disequilibrium (LD) structure, haplotypes and recombination hotspots in several populations from the different ancestral geographic locations (60). With this knowledge, the first genotyping arrays with considerable coverage of common human genetic variation were designed. Thus, from 2005 these arrays began to be used in genome-wide association studies (GWAS), whose aim is the understanding of the genetic component of a disease or trait, as well as candidate gene studies (61). Nevertheless, GWAS allow the analysis of millions of variants distributed throughout the genome being this its main advantage (62). In addition, they are hypothesis-free studies, which allow working without any prior assumptions (63). The methodology commonly applied in GWAS is based on the genotyping of genomic common SNPs in collected individuals with phenotypic data and clinical information for a given trait or disease; after the corresponding quality controls, the genotypic

data can be phased by the imputation process, assessing a larger number of SNPs and increasing the coverage; later, the association test is assess and the associated signals will be those that pass the genome-wide significance threshold, set at *P-value*<$5x10^{-08}$ that corresponds to the Bonferroni correction for 1,000,000 independent SNPs (Figure 4) (63, 64). The results obtained are usually tested in independent cohorts in a replication process, as well as *in silico* analyses are carried out on available sources to check the relevance of the results (63) (Figure 4).

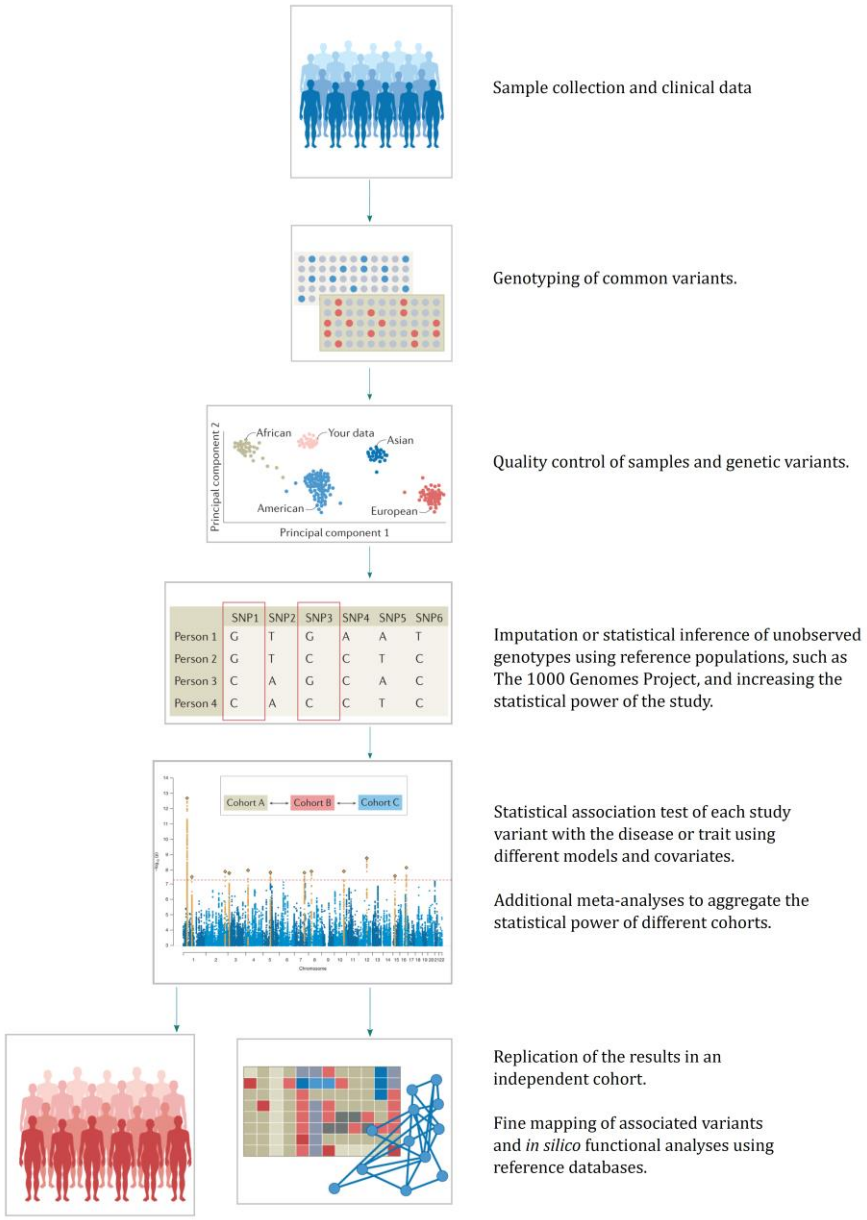

Figure 4. Main steps for genome-wide association (Modified from Uffelmann E, Nat Rev Methods Primers. 2021).

The launch of the 1000 Genomes Project (1KGP) aimed to characterize the human genome and cover low-frequent variants as well as variation with different frequency among populations. This project included 14 representative populations from Europe, East Asia, sub-Saharan Africa and the Americas, supposing the last one an advance in the representation of Latin American populations in reference data (54, 65). In this sense, the inclusion of populations in the genetic studies with mixed ancestry supposed a challenge, as is the case of Latin American populations that have a recently admixture with Native American, European and African continental sources (66). In addition, individuals from these populations show inter-individual variation in ancestry proportions, which has been correlated with the risk of other infectious diseases such as malaria (67, 68). This situation allows the application of admixture mapping studies with the aim to assess disease associations with the varying local ancestries in admixed populations, as a GWAS complementary methodology (69, 70).

Nevertheless, the DNA sequence variations are not the only mechanism that can be associated with a specific phenotype. The epigenetic comprises the set of modifications in gene expression without changing the DNA sequence. These changes can take place under the influence of environmental factors or the genetic variation itself (Figure 5) (71). The most studied epigenetic mechanism is the DNA methylation that consists in a heritable modification of the cytosine DNA by the addition of a methyl group, which in most cases uses to inhibit gene expression (72). Given that the majority of genetic variants are located in non-coding regions, functional studies are required in order to link those associations with specific genes or pathways. Thus, the combination of epigenomic and genomic data allows the assessment of biological and functional significance of the variants associated with a disease. In turn, methylation patterns can be influenced by variations in the DNA sequence, which are known as methylation

quantitative trait *loci* (mQTL) (72). The analysis of these interactions allows us to give functional significance to the associated SNPs, which is essential for the genetic characterization of the disease. Therefore, genetic and epigenetic variations and their interaction can be translated into expression changes. Transcriptomic analyses are a good strategy to evaluate the consequences of genetic variations and their interactions (73). These analyses together with the online tools and databases available for different populations make it possible to evaluate the genetic component of different type of diseases, as well as to functionally evaluate disease-associated variants for unraveling the biological mechanisms underlying associations.

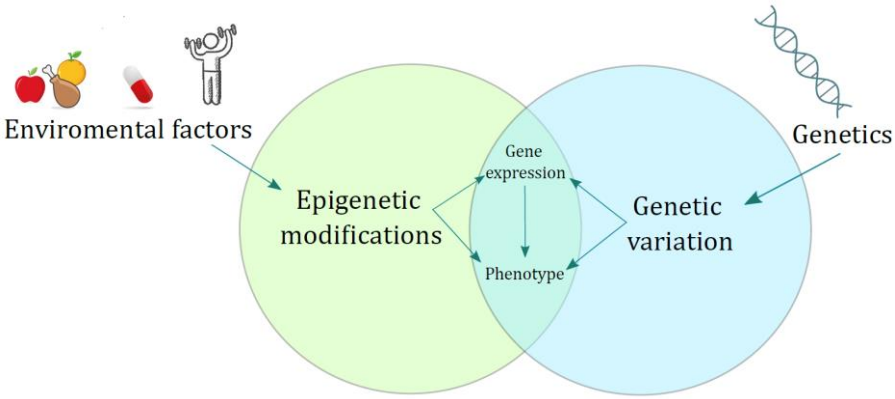

Figure 5. Factors influencing genetic and epigenetic variability and their relationship to gene expression. The dynamic among these modifications can be evaluated in order to assess their relationship with the development or the susceptibility to the disease. (Modified from Arif M, Hypertens Res. 2019).

5. Genetic analysis of infectious diseases

Regarding the host genetic burden in infectious diseases, it is hypothesized that it contributes to the heterogeneity in the response and the

severity of the infection (74). This hypothesis arises from the idea of a continuous exposure of our genomes to the infectious agents for several generations and that, through the action of co-evolution and natural selection, produces this variability in the susceptibility to these diseases (75). In this sense, the evidence of this contribution has been assessed in several infectious diseases, such as malaria, HIV-1 or tuberculosis, mainly through GWAS (74).

One of the most studied parasitic diseases is Malaria. This infections is caused by different *Plasmodium* species and several candidate gene studies and GWAS have reported genetic associations with the development of the disease and its severe forms in several African populations (76-78). Therefore, strong associations have been described for the *ABO*, *CD40L* and *ATP2B4* genes with the disease (79). In the same way, the Leishmaniasis is a parasitic infection caused by species from the genus *Lesinmania* where the large number of molecular studies carried out in mouse models has allowed the identification of important genes and pathways in the disease (79). Based on the last point, most of the genetic studies carried out in this disease have been candidate gene studies, highlighting the *SLC11A1* and *FL1* genes and their association with different forms of the disease (80-82). Recently, a GWAS in cutaneous leishmaniasis has been published in which some suggestive associations in the *LAMP3* and *STX7* genes were identified, being these genes previously related to the intracellular localization of the parasite (83). Regarding other trypanosomiasis, the African trypanosomiasis is caused from another member of the *T. cruzi* family, *T. brucei*. Thus, candidate gene studies highlighted genes related with the immune response, such as *IL6*, or genes related with the identification of the parasite by the host, such as *APOL1* (84).

In this way, the large amount of information available today that confirms the host differential genetic susceptibility to contract different infectious diseases highlights the need to deepen this type of genetic studies in order to translate it to clinical and preventive benefit.

5.1. Previous genetic studies in Chagas disease

Genetic studies are great tools to improve the knowledge about the molecular basis of Chagas disease, opening new avenues for the development of biomarkers of disease progression, new therapeutic strategies to suit the requirements of individual patients, and contributing to the control of one of the infections with the greatest socio-economic impact in the Americas. Thus, several studies have assessed the influence of genetic variation in the susceptibility to Chagas disease and the development of the chronic forms, especially in those genes with implication in the immune response.

The major histocompatibility complex (MHC) region is located on chromosome 6 and encode different sets of proteins that play key roles in the immune system and that are responsible for mediating host-pathogen interactions (85). The most relevant molecules in this context are the human leukocyte antigen (HLA), which are cell-surface glycoproteins that bind peptide fragments that either have been synthesized within the cell (HLA class Ia; A, B, and C) or that have been ingested and processed by the cell (HLAclass II; DR, DQ, and DP). Several HLA alleles and haplotypes have been reported to be associated with the susceptibility to *T. cruzi* infection and to the chronic manifestation of Chagas disease (86, 87) (Table 1).

Table 1. HLA genes

Gene	Alleles/Genotyping	Population	Association	Reference
HLA class I and II				
HLA-A	A*68	Mexico	CCC	(Cunha-Neto et al., 2014)
	A*30	Brazil	All chronic forms	(Deghaide et al., 1998)
HLA-B	B*39	Mexico	Infection and CCC	(Cruz-Robles et al., 2004)
	B*35	Mexico	CCC	(Cruz-Robles et al., 2004)
	B*40 in the presence of Cw3	Chile	CCC	(Llop et al., 1988)
	B*14:02	Bolivia	All chronic forms	(del Puerto et al., 2012)
HLA-C	C*03 in LD with B*40 and B*15	Venezuela	CCC	(Layrisse et al., 2000)
	KIR2DS2 in the presence of HLA-C1	Brazil	CCC	(Ayo et al., 2015b)
HLA-DRB1	DRB1*14-DQB1*0301 haplotype	Peru	Infection	(Nieto et al., 2000)
	DR16 and DR4	Mexico	Infection and CCC	(Cruz-Robles et al., 2004)
	DRB1*1503 and DRB1*1103	Argentina	Infection and CCC	(Garcia Borras et al., 2009)
	DRB1*0409	Argentina	Infection	(Garcia Borras et al., 2006)
	DRB1*01, DRB1*08 and DRB1*1501	Venezuela	CCC	(Fernandez-Mestre et al., 1998)
	DRB1*01 and DQB1*0501	Venezuela	CCC	(Colorado et al., 2000)
	DRB1*01	Bolivia	All chronic forms	(del Puerto et al., 2012)
HLA-DQB1	DQB1*0303	Venezuela	CCC	(Fernandez-Mestre et al., 1998)
	DQB1* 0302 and DQB1*06	Brazil	All chronic forms	(Deghaide et al., 1998)
HLA-DPB1	DPB1* 0401, DPB1*2301and DPB1*3901	Venezuela	CCC	(Colorado et al., 2000)
MICA	MICA MICA*011	Bolivia	Infection	(del Puerto et al., 2012)
	DRB1*01-B*14-MICA*011 haplotype	Bolivia	All chronic forms	(del Puerto et al., 2012)
	MICA-129	Brazil	CCC	(Ayo et al., 2015b)
MICB	MICB*008	Bolivia	No association	(del Puerto et al., 2012)
HLA-G	3' UTR	Brazil	All chronic forms	(Dias et al., 2015)

Table 1. HLA genes

Gene	Alleles/Genotyping	Population	Association	Reference
HLA class III				
TNF	TNFa -238A	Brazil	Infection	(Pissetti et al., 2011)
	TNFa -1031C and -308A	Colombia	CCC	(Criado et al., 2012)
	TNFa-308	Mexico	CCC	(Rodriguez-Perez et al., 2005)
	TNFa microsatellite and -308	Brazil	CCC	(Drigo et al., 2006)
	TNFd3, TNFb7, TNFa8; TNFa2-b4-c2-d3-e2 and TNFa8-b1-c1-d3-e3 haplotypes	Brazil	Infection and All chronic forms	(Campelo et al., 2007)
	TNFa microsatellite and -308	Brazil	No association	(Drigo et al., 2007)
	TNFa -308, -244 and -238 and TNFb	Peru	No association	(Beraun et al., 1998)
	TNFa -308	Brazil	No association	(Alvarado-Arnez et al., 2018)
	TNFa -308	Brazil	No association	(Lima et al., 2018)
LTA	LTA +80 and MAS252	Brazil	CCC	(Ramasawmy et al., 2007)
	LTA +252	Brazil	CCC	(Pissetti et al., 2013)
	LTA +80 and +252	Brazil	No association	(Alvarado-Arnez et al., 2018)
BAT1	BAT1 – 22C/G and –348C/T	Brazil	CCC	(Ramasawmy et al., 2006a, 2006b)
	BAT1 – 22C/G	Brazil	CCC	(Alvarado-Arnez et al., 2018)
CYP21A2	V281L	Bolivia	CCC	(del Puerto et al., 2013)
IKBL	IKBL–62A/T and –262A/G	Brazil	CCC	(Ramasawmy et al., 2008)

The most studied genes are *HLA-B and HLA-DRB1*. A study conducted in Peruvian population found that DRB1*14-DQB1*0301 haplotype was associated to protection against *T. cruzi* infection in a highly endemic area (88). In the same way, another study in the Mexican population found that the frequencies of *HLA-DR4* and *HLAB39* were significantly increased in seropositive individuals when compared to healthy controls, while the frequency of the *HLA-B35* increased in patients with cardiomyopathy when compared to healthy controls (89). A study performed in Bolivian population found that the frequencies of HLA-B*14:02 and HLA-DRB1*01 were significantly lower in patients suffering from megacolon as well as in those with electrocardiogram alteration and/or megacolon compared with a group of patients with indeterminate symptoms (90). Furthermore, the HLA-B*3505 allele was associated with moderate to severe cutaneous reaction in response to Benznidazole (91). Another less polymorphic *HLA-G, MICA* and *MICB* have also been studied (90, 92, 93), where the *HLA-G* gene exhibited distinct patterns of associations with cardiac and digestive chronic Chagas forms in Brazilian population (93, 94). The variability of HLA alleles associated with Chagas disease in the mentioned studies may reflect the genetic heterogeneity presented among the different population in Latin America.

Within the MHC class III region, the *TNF* and *LTA* genes are the most widely investigated regarding the susceptibility to the infection and the chronic cardiac form. Thus, some studies demonstrated that the *TNF* gene polymorphisms could modify the genetic risk of developing heart disease (95, 96). Nevertheless, although genetic associations were reported in Colombian, Mexican and Brazilian populations (97-100), other authors did not find a significant association between the *TNF* gene variants with the disease (101-103).

Cytokines are mediators of the immune response and help to modulate the progression of the disease by inhibiting parasitic replication in different cell types (101). Most candidate gene studies carried out in Chagas disease have been focused on these molecules and were replicated in different Latin American populations (Table 2). *IL-17*, *IL-18* and *IFN-γ* are critical molecules for host defense against a variety of intracellular pathogens. Several studies shown that polymorphisms located in the genes encoding for these cytokines were associated with the susceptibility to *T. cruzi* infection and the chronic Chagas cardiomyopathy, mainly in Colombian and Brazilian populations (101, 104-107). In the case of *IL-18* and *IFNG-*, polymorphisms in these genes have been associated with the risk of infection but not with the development of the chronic cardiac form in Colombian patients (104, 105, 108), while in the case of the Brazilian population associations with the chronic form were reported (101, 109). The authors explained that these discrepancies may be the result of a difference in genotypic distributions between Brazilian and Colombian populations.

Table 2

Cytokines, chemokines and their receptors.

Gene	Chr	Gene name	Population	Association	Reference
Cytokines and their receptors					
IL-1A	2	Interleukin 1 alpha	Colombia	CCC	(Florez et al., 2006)
IL-1B	2	Interleukin 1 beta	Colombia	CCC	(Florez et al., 2006)
			Mexico	No association	(Cruz-Robles et al., 2009)
IL-1RN	2	Interleukin 1 receptor antagonist	Mexico	CCC	(Cruz-Robles et al., 2009)
			Colombia	No association	(Florez et al., 2006)
IL-4	5	Interleukin 4	Bolivia	Infection	(Alvarado Arnez et al., 2011)
			Colombia	No association	(Florez et al., 2011)
IL-4R	16	Interleukin 4 receptor	Colombia	CCC	(Florez et al., 2011)
IL-6	7	Interleukin 6	Colombia/Peru	No association	(Torres et al., 2010a)
			Brazil	CCC	(Costa et al., 2009)
IL-10	1	Interleukin 10	Colombia	No association	(Florez et al., 2011)
			Brazil	No association	(Alvarado-Arnez et al., 2018)
IL-12B	5	Interleukin 12B	Colombia	CCC	(Zafra et al., 2007)
IL-17A	6	Interleukin 17A	Colombia	Infection and CCC	(Leon Rodriguez et al., 2015)
IL-18	11	Interleukin 18	Colombia	Infection	(Leon Rodriguez et al., 2016)
			Brazil	CCC	(Nogueira et al., 2015)
IFNG	12	Interferon gamma	Colombia	Infection	(Torres et al., 2010b)
			Brazil	CCC	(Alvarado-Arnez et al., 2018)
TGFB	19	Transforming growth factor beta	Colombia/Peru	Infection	(Calzada et al., 2009)
			Brazil	No association	(Alvarado-Arnez et al., 2018)

Table 2

Cytokines, chemokines and their receptors.

Gene	Chr	Gene name	Population	Association	Reference
Chemokines and their receptors					
CCR5	3	C-C motif chemokine receptor 5	Peru	CCC	(Calzada et al., 2001a)
			Colombia	CCC	(Florez et al., 2012; Machuca et al., 2014)
			Venezuela	CCC	(Fernandez-Mestre et al., 2004)
			Argentina	CCC	(Juiz et al., 2019)
			Brazil	CCC	(Nogueira et al., 2012)
			Brazil	CCC	(Frade et al., 2013)
			Brazil	All chronic forms	(de Oliveira et al., 2015)
			Brazil	No association	(Lima et al., 2018)
			Colombia	CCC	(Florez et al., 2012; Machuca et al., 2014)
CCR2	3	C-C motif chemokine receptor 2	Argentina	CCC	(Juiz et al., 2019)
CCL2	17	C-C motif chemokine ligand 2	Brazil	CCC	(Frade et al., 2013; Ramasawmy et al., 2009)
CXCL9	4	C-X-C motif chemokine ligand 9	Brazil	CCC	(Nogueira et al., 2012)
CXCL10	4	C-X-C motif chemokine ligand 10	Brazil	CCC	(Nogueira et al., 2012)
MIF	22	Macrophage migration inhibitory factor	Colombia/Peru	Infection	(Torres et al., 2009)
CCL5	17	C-C motif chemokine ligand 5	Brazil	No association	(Nogueira et al., 2012)
CCL17	16	C-C motif chemokine ligand 17	Brazil	No association	(Nogueira et al., 2012)
CCL19	9	C-C motif chemokine ligand 179	Brazil	No association	(Nogueira et al., 2012)

Chemokines have been also directly related to cardiac effects, involving processes such as heart tissue repair, arrhythmia and cardiac insufficiency (87). The chemokine receptor 5 (CCR5) is a receptor for the chemokines CCL2, CCL3, CCL4 and CCL5 and is one of the most replicated gene. Interestingly, patients with cardiomyopathy exhibited higher expression of this gene, which resulted in increased inflammation (110). In addition, CCR2 belongs to the same chemokine receptor family and has also been associated with chronic Chagas cardiomyopathy after *T. cruzi* infection. Thus, *CCR5*, *CCR2* and their haplotypes have been associated with the development of this chronic cardiomyopathy (111, 112). On its part, chronic Chagas cardiomyopathy patients with ventricular dysfunction displayed reduced genotypic frequencies of variants in *CXCL9*, *CXCL10*, and increased in *CCR5* as compared to those without the dysfunction (113). Additionally, the chemokine macrophage migration inhibitory factor (MIF) is an integral component of the host antimicrobial alarm system whose coding gene has been also associated with the infection by *T. cruzi* in Colombian and Peruvian populations (114).

Several other genes, selected for their previous association with other infectious diseases, have been studied for Chagas disease, mainly those related with the immune response. These include genes coding for TLRs and related molecules such as Toll-interleukin-1 receptor domain containing adapter protein (TIRAP). This molecule encodes an adapter protein associated with TLRs, which recognizes microbial pathogens. In the context of Chagas disease two studies have reported an association with the susceptibility of developing chronic Chagas cardiomyopathy (Table 3) (115, 116). Others studies have addressed the association between haptoglobin (HP) polymorphisms with *T. cruzi* infection and its chronic forms, including cardiac and digestive forms in Brazilian and Venezuelan population (117,

118). HP is an acute-phase protein synthesized mainly by the liver during inflammatory processes and also possesses anti-inflammatory and antioxidant properties (119). On the other hand, two studies in the mannose-binding lectin 2 (MBL2) showed association with an increased risk of severe cardiac Chagas disease, related with high MBL serum levels in Brazilian population (120), and a moderate association with the risk of infection in Chilean population (121).

Table 3

Others genes.

Gene	Chr	Gene name	Population	Association	Reference
HP	16	Haptoglobin	Brazil	CCC	(Jorge et al., 2010)
			Venezuela	Infection and CCC	(Mundaray Fernandez et al., 2014)
MBL2	10	Mannose binding lectin 2	Brazil	Infection and CCC	(Luz et al., 2016)
			Chile	Infection	(Weitzel et al., 2012)
TIRAP	11	Toll-interleukin receptor domain containing adaptor protein	Brazil	CCC	(Frade et al., 2013; Ramasawmy et al., 2009)
VDR	12	Vitamin D3 receptor C	olombia	CCC	(Leon Rodriguez et al., 2016)
MASP2	1	Mannan binding lectin serine peptidase 2	Brazil	CCC	(Boldt et al., 2011)
ACTC1	15	Actin alpha cardiac muscle 1	Brazil	CCC	(Frade et al., 2013)
FCN2	9	Ficolin 2	Brazil	All chronic forms	(Luz et al., 2013)
CTLA4	2	Cytotoxic T-lymphocyte associated protein 4	Brazil	CCC	(Dias et al., 2013)
			Venezuela	No association	(Fernandez-Mestre et al., 2009)
TLR1/TLR2	4	Toll like receptor 1 / Toll like receptor 2	Chile	No association	(Weitzel et al., 2012)
			Brazil	No association	(Ramasawmy et al., 2009)
TLR2	4	Toll like receptor 2	Colombia	No association	(Zafra et al., 2008)
TLR4	9	Toll like receptor 4	Colombia	No association	(Zafra et al., 2008)
			Brazil	No association	(Ramasawmy et al., 2009)
TLR5/TLR9	1/3	Toll like receptor 5 / Toll like receptor 9	Brazil	No association	(Ramasawmy et al., 2009)
TLR6	4	Toll like receptor 6	Chile	No association	(Weitzel et al., 2012)
PTPN22	1	Tyrosine-protein phosphatase non-receptor type 22	Colombia/Peru	No association	(Robledo et al., 2007)
NRAMP1	2	Human natural resistance-associated macrophage protein 1	Peru	No association	(Calzada et al., 2001a)
NOS2	17	Nitric oxide synthase	Peru	No association	(Calzada et al., 2002)
FOXO3	6	Forkhead box protein 3	Colombia	No association	(Leon Rodriguez et al., 2016)
TYK2	19	Non-receptor tyrosine-protein kinase	Colombia	No association	(Leon Rodriguez et al., 2018)

OBJECTIVES

Despite previous evidences of the role of the host genetic component in Chagas disease and its chronic cardiac form, the knowledge of the genetic architecture and its functional involvement in the disease is limited. Therefore, the present thesis has the overall objective of determining the genetic basis and molecular mechanisms underlying Chagas disease and its most severe chronic form, the chronic Chagas cardiomyopathy.

The specific objectives were:

1. To assess previously reported genetic associations with Chagas disease using the candidate genes strategy in different Latin American populations.

2. To perform a GWAS in Chagas disease to identify new *loci* associated with susceptibility to infection and the development of chronic Chagas cardiomyopathy.

3. To explore the influence of genetic ancestry in relation to *T.cruzi* infection using an admixture mapping approach.

4. To functionally characterize the GWAS-*loci* associated with chronic Chagas cardiomyopathy by integrating genomic data with DNA methylation data.

MATERIAL AND METHODS, RESULTS AND DISCUSSION

Chapter 1: Genetic polymorphisms of *IL6*, *IL17A* and *IL18* genes and their associations with Chagas disease.

As mentioned in the previous chapter, genetic factors and the immune response have been suggested to be determinant in the susceptibility against the infection and the outcome of Chagas disease. For this, the analysis of genetic variants localized in the vicinity of genes involved in the immune response, such as the interleukin encoding genes, might give novel insights in the pathogenesis of the infection and its chronic cardiac form. In this chapter we analyzed *IL6*, *IL17A* and *IL18* genetic variants regarding the predisposition to *T. cruzi* infection and the development of chronic Chagas cardiomyopathy in different Latin American populations. Individuals were classified as seronegative and seropositive for *T. cruzi* antigens, and the last group was divided into asymptomatic and chronic Chagas cardiomyopathy patients. The results showed association of the *IL17A* and *IL18* genetic variants with the differential susceptibility to the infection, while only the *IL18* polymorphisms showed association with the chronic form of the disease, confirming previously published results.

1.1. Materials and methods

1.1.1. Ethics statement

The study was accepted by the Ethics Committees from the Industrial University of Santander and Cardiovascular Foundation, Colombia (Act No. 15/2005), the Vall d'Hebron University Hospital, Barcelona, Spain (PR (AMI) 297/2016), and the National Hospital of Clinics, National University of Cordoba, Argentina (CIEIS HNC 118/2012 and 2/16/2017). Written informed consent was obtained from all subjects prior to participation. The research protocols followed the principles of the Declaration of Helsinki, and informed consent was obtained from all individual participants included in the study.

1.1.2. Study design and patient populations

Candidate-gene case-control studies were performed in Colombian, Bolivian and Argentinian cohorts in order to replicate previous findings and test the association of *IL6, IL17A* and *IL18* polymorphisms. Additionally, a meta-analysis was performed combining the tested cohorts for each analyzed gene. Sample sized of each cohort for the *IL6*, *IL17A* and *IL18* genes are summarized in Table 1.1.

Table 1.1. Sample size from the different cohorts included in each candidate gene study for *T. cruzi* infection and the chronic Chagas cardiomyopathy.

Gene		Colombia	Bolivia	Argentina	Brasil
IL6	Sex (%female)	55%	69%	71%	-
	Age (mean±SD)	52.12±17.65	48.83±9.45	53.82±13.52	-
	n (seropositive/seronegative)	744/647	-	272/78	-
	n (CCC/asymptomatic)	468/276	100/530	182/90	-
IL17A	Sex (%female)	58%	69%	71%	51.71%
	Age (mean±SD)	45.55±17.19	48.83±9.45	53.82±13.52	62.60±13.70
	n (seropositive/seronegative)	937/640	-	272/78	260/150
	n (CCC/asymptomatic)	576/361	100/530	182/90	212/48
IL18	Sex (%female)	58%	69%	71%	52%
	Age (mean±SD)	45.55±17.19	48.83±9.45	53.82±13.52	-
	n (seropositive/seronegative)	937/640	-	272/78	-
	n (CCC/asymptomatic)	576/361	100/530	182/90	849/202

Total of individuals included in each candidate gene study from the different populations. CCC chronic Chagas cardiomyopathy.

Colombian cohort: Colombian samples were recruited in an endemic region for Chagas disease (Guanentina and Comunera provinces, Santander department) by the health care team from the Industrial University of Santander and Cardiovascular Foundation from Colombia. Samples were classified as seronegative or seropositive according to the serological tests of recombinant antigen enzyme-linked immunosorbent Assay (ELISA) and commercial indirect hemagglutination. Additionally, basing on complementary tests and clinical findings, seropositive patients were classified as chronic Chagas cardiomyopathy patients or asymptomatic individuals. Thus, in the case of the study of the *IL6* polymorphisms, a total of 1,391 Colombian individuals were assessed for the susceptibility to the infection and the development of the chronic cardiac form of the disease. The genetic data for the selected SNPs for this gene of these individuals were additionally meta-analyzed with previously published data from Colombia (Colombia I) and Peru (107). The mean age of the participants was 52.12±17.65 years. Regarding the study of *IL17A* and *IL18* polymorphisms, a total of 406 Colombian

individuals from the same population as the study by Leon Rodriguez et al. (2016) were taking into account for the analyses (122). In order to increase the sample size, these individuals were included with the previously published Colombian cohort, making a total of 1,577 individuals to assess the susceptibility to both infection and chronic form.

Bolivian cohort: A total of 630 Bolivian individuals residents in Barcelona, Spain were recruited from the Infectious Diseases Department of the Vall d'Hebron University Hospital. This cohort is composed only by seropositive individuals, which were subjected to electrocardiograms, echocardiograms, and chest radiography in order to identify cardiac involvement. The mean age of the chronic group was 50.71±9.41, while for the asymptomatic individuals was 46.93±9.49.

Argentinian cohort: A total of 350 Argentinian individuals from an endemic region for Chagas disease (Cordoba province) were included in this study. The samples were recruited from the National Hospital of Clinics and Sucre Clinic, Cordoba city. As for the Colombian cohort, all participants underwent a serological diagnosis for *T. cruzi* infection through ELISA assays in combination with a commercial indirect hemagglutination test. Seropositive individuals were classified as classified as chronic Chagas cardiomyopathy patients and asymptomatic according to the results of complementary tests and clinical findings, seropositive individuals were. The mean age of participants was 53.82±16.53 years for seronegative individuals, 60.14±10.16 for chronic patients and 49.30±13.65 for the asymptomatic group.

Brazilian cohort: The Brazilian population was included in the study of *IL17A* and *IL18* polymorphisms based on previous studies. Regarding the study of *IL17A* polymorphisms, data from 410 Brazilian individuals drawn from Reis et al. 2017, originally from South and Southeast regions of Brazil were included in the meta-analysis for the susceptibility to the infection and the chronic Chagas cardiomyopathy (123). In the case of *IL18*, a total of 1,051 Brazilian seropositive patients for antibodies against *T. cruzi* were included in the meta-analysis.

1.1.3. Selected polymorphisms and genotyping.

For each gene under study, selected polymorphisms corresponded to: rs1800795, as a promoter variant of the *IL6* gene; rs4711998, rs8193036 and rs2275913 for the *IL17A* gene, previously assessed in Chagas disease in a Colombian cohort and in a Brazilian cohort (123); and the rs2043055, rs1946518 and rs360719 SNPs in representation of the *IL18* gene. LD (R^2 and D') was estimated using an expectation–maximization algorithm implemented in Haploview V4.2 (124) for the studied cohorts and from the American sub-populations genotype data from the 1KGP phase III (54) in the case of the *IL18* polymorphisms in conjunction with the LDlink website tool (https://ldlink.nci.nih.gov/?tab=ldmatrix).

Genomic DNA from blood samples was isolated following standard procedures, and the genotyping was performed using TaqMan assays (Applied Biosystems, Foster City, California, USA) on a real-time PCR system (7900HT Fast Real-Time PCR System), SNPs were determined by TaqMan 5´ allelic discrimination assay method performed by Applied Biosystems.

1.1.4. Statistical analyses

For the candidate gene study, the statistical analyses were performed with the software Plink V1.9 (125). Deviance from Hardy-Weinberg equilibrium (HWE) was determined at the 1% significance level in all groups of individuals. Individuals that did not achieved an SNP completion rate of 95% were filtered out. To test for possible allelic association, logistic regression model and Fisher's exact test were assessed in seropositive vs. seronegative individuals and chronic Chagas cardiomyopathy individuals vs. asymptomatic, using age and sex as covariates. Odds ratios (OR) and 95% confidence intervals (CI) were calculated and in those cases that control for multiple testing were necessary the Benjamini & Hochberg step-up false discovery rate (FDR) correction was used. *P-values* lower than 0.05 were considered as statistically significant.

To assess the consistency of effects across the cohorts, a meta-analysis was performed with METASOFT (126) based on inverse-variance-weighted effect size. Heterogeneity across studies was assessed using the Cochran's Q statistic (Q test *P-value*<0.05) and I^2 heterogeneity index (126). A fixed-effects model was applied for those SNPs without evidence of heterogeneity (Cochran's Q test *P-value*>0.05), and a random-effects model was applied for SNPs displaying heterogeneity of effects between studies (Cochran's Q test *P-value*≤0.05). The statistical power of the studies was estimated with the Power Calculator for Genetic Studies 2006 (CaTS) software (http://www.sph.umich.edu/csg/abecasis/CaTS) (127) and the allele frequency used for statistical power calculation was the one described for the Americans sub-populations of the 1KGP phase III (54). Evaluation of functionality of the SNPs was performed with the online software

HaploReg v4.1 (128) based on empirical data from the ENCODE project (http://www.genome.gov/encode) and focused specifically on experiments performed on blood and T cells lines in the American population. For regulatory features, Ensembl Browser (54) and ReMap 2018 v1.2 (129) were used.

1.2. **Results**

The variants under study were in HWE in all the analyzed cohorts (*P-value*>0.01) and the genotyping success rate was >90% for all of them. In addition, allele frequencies in all cases were similar to those described for the Americans sub-populations of the 1KGP phase III (54). Demographic information and sample sizes for the included cohorts are summarized in Table 1.

1.2.1. IL6 results

Regarding the association of *IL6* with *T. cruzi* infection susceptibility, allelic and genotypic frequencies of seronegative and seropositive individuals from the Colombian and Argentinian cohorts were compared for the variant rs1800795 (Table 1.2). In both cohorts, the frequency of the minor allele C, was increased in seronegative compared with seropositive individuals, but no significant association between *IL16* rs1800795 and susceptibility to *T. cruzi* infection was found after applying logistic regression adjusted by sex and age. In the meta-analysis, when combining the newly recruited cohorts (Colombia and Argentina) and the previously published data (Colombian I and Peruvian), the effect of the C allele was consistent among the populations and indicated protection against *T. cruzi* infection. However, no statistically significant differences were observed under a fixed-effects meta-analysis (*P-value*= 0.433, OR=0.94, 95%CI= 0.81-1.10) (Table 1.3).

Table 1.2. Genotype and allele distribution of the *IL6* genetic variant rs1800795 in seronegative vs. seropositive and CCC vs. Asymptomatic individuals in the Colombian, Bolivian and Argentinian cohorts.

			Genotype. N (%)						
Cohorts	Group	Sample size	CC	CG	GG	MAF	OR	(L95-U95)	*P-value*
Seropositive vs. Seronegative									
Colombia	Seronegative	639	41 (6.36)	198 (30.7)	406 (62.95)	21.71	0.91	(0.75-1.12)	0.371
	Seropositive	744	26 (3.51)	247 (33.21)	470 (63.18)	20.1			
Argentina	Seronegative	76	6 (7.79)	22 (28.57)	49 (63.64)	22.08	0.86	(0.56-1.32)	0.479
	Seropositive	272	10 (3.7)	85 (31.48)	175 (64.81)	19.44			
CCC patients vs. Asymptomatic									
Colombia	Asymptomatic	276	6 (2.17)	95 (34.42)	175 (63.41)	19.38	1.12	(0.81-1.54)	0.502
	CCC	468	21 (4.49)	153 (32.69)	294 (62.82)	20.83			
Bolivia	Asymptomatic	529	5 (0.95)	80 (15.12)	444 (83.93)	8.51	1.53	(0.89-2.61)	0.119
	CCC	100	2 (2.00)	18 (18.00)	80 (80.00)	11			
Argentina	Asymptomatic	89	1 (1.12)	32 (35.96)	56 (62.92)	19.1	1.05	(0.62-1.75)	0.868
	CCC	182	9 (4.97)	53 (29.28)	119 (6575)	19.61			

1 minor allele | 2 major allele; alleles are shown in the forward strand. CCC chronnic Chagas cardiomyopathy, MAF minor allele frequency, OR odds ratio, L95-U95 confidence intervals of 95% L lower limit, U upper limit.

In the case of chronic Chagas cardiomyopathy, the rs1800795*C allelic frequency in Colombian, Bolivian and Argentinian cohorts was increased patients compared with asymptomatic individuals, suggesting a risk for Chagas cardiomyopathy susceptibility in these populations, but in none of these cohorts the results were statistically significant after performing logistic regression adjusted by sex and age (Table 1.2). Moreover, although the allele effect was consistent in all the Latin American cohorts analyzed no significant association was detected for *IL6* rs1800795 and chronic Chagas cardiomyopathy when the fixed-effect meta-analysis was performed (Table 1.3) (*P-value*=0.257, OR=1.11, CI=0.93-1.33).

Table 1.3. Meta-analysis of *IL6* rs1800795*C in *T. cruzi* infection susceptibility and the development of chronic Chagas cardiomyopathy in Latin American cohorts.

Seropositive vs. seronegative									
Colombia (N=1,383)		Colombia I (N=554)		Argentina (N=348)		Peru (N=162)		Meta-analysis (N=2,447)	
OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*
0.91 (0.74-1.12)	0.371	1.11 (0.84-1.48)	0.466	0.86 (0.56-1.32)	0.479	0.55 (0.23-1.31)	0.183	0.94 (0.81-1.10)	0.433
CCC patients vs. Asymptomatic									
Colombia (N=744)		Colombia I (N=239)		Argentina (N=271)		Peru (N=78)		Bolivia (N=629)		Meta-analysis (N=1,961)	
OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*
1.12 (0.81-1.58)	0.502	1.12 (0.72-1.70)	0.651	1.05 (0.62-1.75)	0.868	3.17 (0.8-12.63)	0.107	1.53 (0.89-2.61)	0.119	1.11 (0.93-1.33)	0.257

CCC chornic Chagas cardiomyopathy, N total individuals included from each cohort, OR odds ratio, L95-U95 confidence intervals of 95% L lower limit; U upper limit.

1.2.2. IL17A results

In the susceptibility to *T. cruzi* infection analysis, only two *IL17A* genetic variants were significant when comparing seropositive and seronegative individuals. Therefore, those variants significantly different between cases and controls after adjust by sex and age, were rs8193036*C and rs2275913*A for the Colombian and Brazilian cohorts, respectively (Table 1.4). In the case of the Argentinian cohort no associations of the variants tested was found. Furthermore, the meta-analysis combining each individual cohort (Colombian, Argentinian and Brazilian) for *IL17A* was performed (Table 1.4). The rs2275913*A allele effect was consistent in the three cohorts and the association improved after the meta-analysis, showing statistically significant results (*P-value*=0.016, OR=1.21, 95% CI=1.06–1.41, under a fixed-effects meta-analysis) after Bonferroni correction, however no association was observed for the *IL17A* rs4711998 and rs8193036 variants in the meta-analysis.

Table 1.4. Meta-analysis of *IL17A* variants, Latin American cohorts for *T. cruzi* infection susceptibility (seropositive vs. seronegative individuals) and for chronic Chagas cardiomyopathy development (CCC vs. asymptomatic patients and CCC vs. seronegative individuals).

Seropositive vs. Seronegative								
	Colombia		Argentina		Brazil		Meta-analysis	
SNP	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*
rs4711998*A	0.94 (0.78–1.14)	0.582	1.38 (0.90–2.12)	0.143	-	-	0.99 (0.84–1.17)	0.946
rs8193036*C	0.83 (0.70–0.99)	0.043	1.34 (0.89–20.2)	0.164	-	-	0.89 (0.76–1.05)	0.169
rs2275913*A	1.16 (0.95–1.4)	0.136	1.07 (0.67–1.69)	0.793	1.46 (1.05–2.05)	0.032	1.21 (1.06-1.41)	0.016

CCC patients vs. Asymptomatic										
	Colombia		Argentina		Bolivia		Brazil		Meta-analysis	
SNP	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*
rs4711998*A	0.86 (0.67–1.11)	0.259	1.08 (0.69–1.68)	0.751	0.96 (0.65–1.41)	0.831	-	-	0.92 (0.76–1.11)	0.396
rs8193036*C	0.92 (0.72–1.18)	0.526	0.74 (0.49–1.29)	0.164	1.18 (0.85–1.62)	0.319	-	-	0.95 (0.80–1.14)	0.616
rs2275913*A	0.8 (0.62–1.02)	0.081	0.72 (0.43–1.21)	0.217	1.14 (0.75–1.71)	0.543	1.21 (0.74–1.99)	0.463	0.89 (0.74–1.07)	0.232

CCC patients vs. Seronegative								
	Colombia		Argentina		Brazil		Meta-analysis	
SNP	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*	OR (L95-U95)	*P-value*
rs4711998*A	0.93 (0.75–1.16)	0.541	1.47 (0.92–2.37)	0.109	-	-	1.02 (0.82–1.23)	0.927
rs8193036*C	0.84 (0.67–1.05)	0.133	1.21 (0.78–1.88)	0.389	-	-	0.91 (0.74–1.10)	0.323
rs2275913*A	1.14 (0.90–1.44)	0.298	0.99 (0.60–1.61)	0.955	1.52 (1.08–2.15)	0.021	1.20 (1.01–1.45)	0.040

OR odds ratio, L95-U95 confidence intervals of 95% L lower limit; U upper limit.

In the case of the analysis of *IL17A* variants in chronic Chagas cardiomyopathy patients, no significant associations were detected in the Colombian, Bolivian and Argentinian cohorts after the logistic regression adjusted by sex and age (Table 1.4). These results were consistent with previous findings in the Brazilian cohort (123). As consequence, any significant association was detected for the available SNPs when the meta-analysis was performed combining these cohorts (Table 1.4). In addition, in order to evaluate the possible association between *IL17A* genetic variants and chronic Chagas cardiomyopathy, CCC patients were compared with seronegative individuals, as previously performed (106, 123). While in the Colombian and Argentinian cohorts no associations were found after applying logistic regression adjusted by sex and age, in the rs2275913*A allele was nominally significant in the Brazilian cohort, as previously reported (123). The effect of this variant is consistent in the Colombian, Argentinian and Brazilian cohorts, and the association with chronic cardiomyopathy susceptibility improved after the meta-analysis showing nominally statistical differences (*P-value*=0.040, OR=1.20, 95% CI=1.01–1.45, under a fixed-effects meta-analysis, Table 1.4).

Given its statistical association with the infection and the chronic phase, an *in silico* functional analysis of the *IL17A* variant rs2275913 and the ones in high LD ($R^2 \geq 0.8$) was performed on peripheral mononuclear blood in American population from the 1KGP (Table 5). The annotation indicates that these SNPs map in enhancer regions and marks of histone modifications (H3K4me1, H3K4me3, H3K27ac and chromatin marks), potentially modulating gene expression.

Table 1.5. Functional annotation. Regulatory chromatin states and histone modifications for IL17A rs2275913 and SNPs in high LD (R^2≥0.8). Functional annotation from mononuclear peripheral blood specifically primary T helper 17 cells.

Position[a]	SNPs	R^2	Functionality	MAF (AMR)	eQTL	Chromatin states[b]	Chromatin states[c]	H3K4me1	H3K4me3	H3K27ac
chr6:52051033	rs2275913	1	Intergenic variant	25%	—	Flank	Promoter	Enhancer	Promoter	Enhancer
chr6:52087034	rs11966760	0.82	Intergenic variant	24%	*PAQR8*	Enhancer	Promoter	Enhancer	Promoter	Enhancer
chr6:52056386	rs16882180	0.8	Intergenic variant	25%	*PAQR8*	—	—	Enhancer	Promoter	-

[a]According to National Center for Biotechnology Genome Reference Consortium NCBI build GRCh37. [b]Core 15-state model. [c]Chromatin states: 25-state model using 12 imputed marks. H3K4me1 Histone H3 lysine 4 mono-methylation, H3K4me3 Histone H3 lysine 4 tri-methylation, H3K27acHistone H3 lysine 27 acetylation, MAF Minor Allele Frequency, AMR American.

1.2.3. IL18 results

In order to assess the association of *IL18* genetic variants with the differential susceptibility to *T. cruzi* infection, the allelic and genotypic frequencies of seronegative and seropositive individuals from Colombia were compared (Table 1.6). The allelic frequencies of the three SNPs were statistically significant even after multiple testing corrections. The frequency of the minor allele, G, in rs2043055 was significantly reduced in the seronegative compared to seropositive individuals suggesting an association with higher infection risk, while the frequencies of rs1946518*T and rs360719*G alleles were significantly increased in seronegative compared to the seropositive individuals, suggesting an association with the protection against the infection by *T. cruzi*. Regarding the Argentinian cohort, no associations between *IL18* genetic variants were found (Table 1.6). However, the rs2043055 remained borderline significant for protection against infection by *T. cruzi* (*P-value*=0.061, OR=0.71, 95% CI=0.49–1.02). In addition, a meta-analysis combining data from Colombian and Argentinean cohorts was performed (Table 1.7). The *IL18* rs360719 showed consistent effects among the two meta-analyzed populations with a statistically significant association (*P-value*= 0.001, 95% CI=0.66–0.89, under a fixed-effects meta-analysis) with an OR for the G allele of 0.76. For this comparison, the sample size attained a statistical power of over 80% for this OR. In both cohorts, the allele effects size were in concordance and this result indicates an association to the protection against *T. cruzi* infection in these cohorts.

Table 1.6. Genotype and allele distribution for *IL18* variants in seronegative and seropositive individuals from the Colombian cohort.

SNP		A1|A2	Genotype N (%)			MAF	Allele test		
			1|1	1|2	2|2		OR	(L95-U95)	P-value
Colombia									
rs2043055	Seronegative	G|A	82(13.00)	300(47.54)	249(39.46)	36.77	1.3	(1.10–1.53)	0.002
	Seropositive		164(17.69)	450(48.54)	313(33.76)	41.96			
rs1946518	Seronegative	T|G	163(25.83)	334(52.93)	134(21.24)	52.3	0.79	(0.67–0.92)	0.003
	Seropositive		214(23.09)	448(48.33)	265(28.59)	47.25			
rs360719	Seronegative	G|A	103(16.32)	299(47.39)	229(36.29)	40.02	0.75	(0.63–0.89)	0.001
	Seropositive		107(11.54)	426(45.95)	394(42.50)	34.52			
Argentina									
rs2043055	Seronegative	G|A	20(25.97)	29(37.66)	28(36.36)	44.81	0.71	(0.49–1.02)	0.061
	Seropositive		33(12.23)	126 (46.67)	111(41.11)	35.56			
rs1946518	Seronegative	T|G	19(24.67)	35(45.46)	23(29.87)	47.4	1.03	(0.71–1.49)	0.883
	Seropositive		54(20.00)	151(55.92)	65(24.08)	47.96			
rs360719	Seronegative	G|A	11(14.28)	33(42.85)	33(42.85)	35.71	0.87	(0.60–1.31)	0.552
	Seropositive		25(9.26)	128(47.40)	117(43.33)	32.96			

1 minor allele | 2 major allele; alleles are showed in forward strand. MAF minor allele frequency, OR odds ratio, L95-U95 confidence intervals of 95% L lower limit, U upper limit. Values adjusted by sex and age. Individuals included in Colombia: 927 seropositive and 631 seronegative, and in Argentina: 270 seropositive and 77 seronegative.

Table 1.7. Meta-analysis of IL18 variants, Latin American cohorts for *T. cruzi* infection susceptibility and Chagas cardiomyopathy development.

	Colombian cohort			Argentinian cohort			Bolivian			Brazilian			Meta-analysis		
SNP	OR	(L95-U95)	*P-value*	OR	(L95-U95)	*P-value*	OR	(L95-U95)	*P-value*	OR	(L95-U95)	*P-value*	OR	(L95-U95)	*P-value*
T. cruzi infection susceptibility															
rs2043055	1.3	(1.10–1.53)	0.002	0.71	(0.49–1.02)	0.061	-	-	-	-	-	-	1.17	(1.01–1.36)	0.035
rs1946518	0.79	(0.67–0.92)	0.003	1.03	(0.71–1.49)	0.883	-	-	-	-	-	-	0.82	(0.71–0.94)	0.006
rs360719	0.75	(0.63–0.89)	0.001	0.87	(0.60–1.31)	0.552	-	-	-	-	-	-	0.76	(0.66–0.89)	0.001
Chronic Chagas cardiomyopathy susceptibility															
rs2043055	0.79	(0.64–0.99)	0.037	1.26	(0.82–1.95)	0.291	1.39	(0.95–2.02)	0.088	1.06	(0.85–1.32)	0.598	1.05	(0.82–1.35)	0.259
rs1946518	1.14	(0.92–1.41)	0.225	0.67	(0.44–1.04)	0.078	1.24	(0.85–1.80)	0.26	-	-		1.07	(0.90–1.26)	0.426
rs360719	0.99	(0.79–1.26)	0.994	0.81	(0.52–1.27)	0.364	0.98	(0.66–1.45)	0.934	-	-		0.95	(0.79–1.15)	0.629

Total number of individuals: seropositive: 1,209 and seronegative: 718, rs2043055 CCC: 1,707 and asymptomatic: 1,183, rs1946518 and rs360719 CCC: 858 and asymptomatic: 981. OR odds ratios, L95-U95 confidence intervals of 95% L lower limit, U upper limit.

The allelic and genotypic frequencies of chronic Chagas cardiomyopathy and asymptomatic patients from Colombia were compared for the *IL18* genetic variant rs2043055, which was statistically significant even after multiple testing correction (*P-value*=0.037, OR=0.79, 95% CI=0.64–0.99, Table 1.8). The frequency of the rs2043055*G allele was significantly incremented in asymptomatic patients, suggesting an association with the protection against the development of Chagas cardiomyopathy. However, no significant differences in allelic frequencies were observed for rs1946518 and rs360719. The rs2043055 was additionally studied in 1,051 seropositive Brazilian patients (Table 1.1) (109). Nevertheless, no significant differences were found in the Brazilian, Bolivian and Argentinian cohorts, only a trend of association can be observed for the SNP rs2043055 in the Bolivian cohort (Table 1.8). On the other hand, the rs1946518 showed an increase of the T allele frequency in asymptomatic individuals from the Argentinian cohort remaining borderline significant (Table 1.8). Further, a meta-analysis combining these results were performed, but the results of the available SNPs showed no significant associations (Table 1.7).

Table 1.8. Genotype and allele distribution for *IL18* variants in asymptomatic and chronic Chagas cardiomyopathy (CCC) individuals from the Colombian, Bolivian and Argentinian cohorts.

SNP		A1|A2	Genotype N (%)			MAF	Allele test		
			1|1	1|2	2|2		OR	(L95-U95)	P-value
Colombia									
rs2043055	Asymptomatic	G|A	83(23.18)	159(44.41)	116(32.40)	45.39	0.79	(0.64–0.99)	0.037
	CCC		81(14.24)	291(51.14)	197(34.62)	39.81			
rs1946518	Asymptomatic	T|G	82(22.91)	160(44.69)	116(32.40)	45.25	1.14	(0.92–1.41)	0.225
	CCC		132(23.20)	288(50.62)	149(26.19)	48.51			
rs360719	Asymptomatic	G|A	45(12.57)	155(43.30)	158(44.13)	34.22	0.99	(0.79–1.26)	0.994
	CCC		62(10.90)	271(47.63)	236(41.48)	34.71			
Bolivia									
rs2043055	Asymptomatic	G|A	72(13.64)	260(49.24)	196(37.12)	38.26	1.39	(0.95–2.02)	0.088
	CCC		16(16.00)	50(50.00)	34(34.00)	41			
rs1946518	Asymptomatic	G|T	101(19.13)	268(50.76)	159(30.11)	44.51	1.24	(0.85–1.80)	0.26
	CCC		19(19.00)	55(55.00)	26(26.00)	46.5			
rs360719	Asymptomatic	G|A	54(10.23)	237(44.89)	237(44.89)	32.67	0.98	(0.66–1.45)	0.934
	CCC		10(10.00)	50(50.00)	40(40.00)	35			
Argentina									
rs2043055	Asymptomatic	G|A	9(10.11)	40(44.94)	40(44.94)	32.58	1.26	(0.82–1.95)	0.291
	CCC		24(13.26)	86(47.51)	71(39.22)	37.02			
rs1946518	Asymptomatic	T|G	21(23.59)	51(57.30)	17(19.10)	52.25	0.67	(0.44–1.04)	0.078
	CCC		33(18.23)	100(55.25)	48(26.51)	45.86			
rs360719	Asymptomatic	G|A	8(8.99)	47(52.81)	34(38.20)	35.39	0.81	(0.52–1.27)	0.364
	CCC		17(9.32)	81(44.75)	83(45.86)	31.77			

Total number of individuals: 569 CCC and 358 asymptomatic in the Colombian cohort; 100 CC and 528 asymptomatic in the Bolivian cohort; and 181 CC and 89 asymptomatic in the Argentinian cohort. 1 minor allele | 2 major allele, alleles are showed in forward strand. MAF minor allele frequency, OR odds ratio, L95-U95 confidence intervals of 95% L lower limit, U upper limit. Values adjusted by sex and age.

The functional annotations of the three *IL18* variants were also explored using HaploReg v4.1, which indicates a location of these variants in a regulatory region of the genome. The annotation based on the epigenomic information of rs2043055 indicates that this SNP maps in an enhancer region, which is correlated with active gene expression in primary mononuclear cells and in T cells from peripheral blood. The rs1946518 and rs360719 variants mapped in a region enriched in histone marks: H3K4me3, H3K9ac, a hallmark of active promoter region, and H3K27ac in enhancer region in mononuclear cells and T cells. Furthermore, according to ReMap 2018 v1.2 these three SNPs mapped in regulatory regions of the human genome and it has been described as transcription factors.

1.3. Discussion

Association studies offer a potentially powerful approach to identify genetic variations that are involved in the immunopathogenesis of Chagas disease (36, 130). However, individual genetic association studies frequently have limitations and the results may be specific to the population of the study. The meta-analysis approach has been proposed to resolve these limitations, to increase the power of statistical analyses (131, 132) and to reach to more conclusive results in order to improve our understanding of the genetic basis underlying Chagas disease.

1.3.1. IL6

Several studies have found associations between *IL6* polymorphisms and infectious diseases susceptibility (133-136). In this sense, the *IL6* polymorphism, rs1800795, has been associated with other affections, such as *Toxoplasma gondii* (134), *Chlamydia trachomatis*, and *Chlamidia pneumoniae* infections (136). In this sense, certain functional polymorphisms in genes encoding cytokines may determine the pathogenesis of Chagas disease, as IL-6 is involved in in the acute and chronic phases of Chagas disease (36, 87, 137-141). For this, several polymorphisms within the *IL6* gene region were assessed to confirm its association with Chagas disease. One of these polymorphisms was the variant rs1800795, which is located in the promoter region of the *IL6* gene so it can modulate its expression (142-144). Although previous studies determined differences in IL-6 expression among *T. cruzi* infected and non-infected patients in previous studies (145, 146), the association analysis of this variant did not show significant associations with the infection or the chronic phase in the five independent Latin American populations. The observed differences in IL-6 expression among *T. cruzi* infected and non-infected patients in previous studies

(145, 146), could be due to wide and complex mechanisms, since its regulation involves interactions between transcription factors, other promoter polymorphisms and DNA methylation (147-151). In spite of this, the lack of association of *IL6* promoter variant rs1800795 in this analysis does not exclude the contribution of other genetic variants within the gene region to be involved in the human susceptibility to Chagas disease.

This is the largest candidate gene study carried out in *IL6*, with a total of 3,087 individuals from different countries included. The statistical power in the meta-analysis was 96% to detect associations with OR=1.3 at the 5% significance level. This, together with the fact that the minor allele frequencies of the variant analyzed presented variation among cohorts, reflect the complex genetic architecture of Latin Americans, a recently admixed population (66, 152). Taking this into consideration, these results do not provide an evidence of significant association between *IL6* rs1800795 with *T. cruzi* infection or chronic Chagas cardiomyopathy in the populations under study. Thus, further studies in other genetic variants of this gene in well-powered cohorts are needed to definitively discard *IL6* gene variants as susceptibility markers for Chagas disease.

1.3.1. IL17A

In the early stages of the infection, the IL-17A is a crucial cytokine secreted by a wide range cell types such as Th17, B cells, innate lymphoid cells, CD4+, CD8+, gamma-delta T and invariant NKT (153-157). Thus, given the importance of this cytokine in the immune response, three *IL17A* genetic variants were assessed in four Latin American populations, Colombia, Argentina, Bolivia and Brazil. The

analyses showed evidence of the implication of rs2275913 in the *T. cruzi* infection and the development of chronic Chagas cardiomyopathy. The rs2275913 is a functional polymorphism that modifies the binding of the transcriptional nuclear factor of activated T cells (NFAT) in the IL-17A promoter and, as showed in the *in silico* analysis, is also located in promoter histone marks, which potentially modulates gene expression. Thus, the association observed in the analysis might be suggesting a variation in gene expression that ends in a lower IL-17A production. In this sense, the reduction in the levels of this cytokine has been described to impede a rapid proinflammatory activation of chemokines and cytokines for the resolution of *T. cruzi* infection (155, 156). On their side, the substitution of the G by an A allele in the associated variant has been demonstrated to be significantly associated with autoimmune diseases and cancer (158-161). However, results reported in serum showed that the A allele was associated with a higher (161-163), lower (164, 165) or no significant (166) levels of transcription and synthesis of the protein. As the results show, the A allele of the rs2275913 shows a risk effect in terms of *T. cruzi* infection. This genetic variation might produce changes in *IL17A* gene expression modifying IL-17A production, which would impede a rapid proinflammatory activation of chemokines and cytokines for the resolution of *T. cruzi* infection (155, 156). To confirm this, functional studies of this variant are required the complexity of the functional effect of this polymorphism.

Several studies showed the immunomodulatory role of IL-17A in the chronic phase of Chagas disease (167-170). *IL17* expression by Th17 cells and B cells were found in patients with cardiac involvement with higher frequency when compared with asymptomatic patients, correlating with worse cardiac function (168). Furthermore, an

exacerbated production of IL-17A produce a proinflammatory environment in Chagas severe heart disease (171, 172). Nevertheless, the present association test showed no significant differences when *IL17A* genetic variants were compared among chronic cardiac patients and asymptomatic individuals from Colombia, Argentina, Bolivia and Brazil were compared. The lack of association observed of *IL17A* genetic variants in the presented analyses might be consequence of an insufficient statistical power or genetic heterogeneity among the populations under study. This lack of replication may occur if the studied polymorphism is not the causal variant but is rather in LD with it, as LD patterns depend on the genetic background of the founder population and population history. In this sense, Latin American populations suppose always a challenge due to their genetic diversity and recent admixture (65, 152, 173). Interestingly, at the present time, IL-17A have become a relevant drug target in various forms of autoimmune and inflammatory diseases, mainly as negative modulators of the secreted protein (156, 174). Thus, two antibodies are currently in Phase IV of drug development for the treatment of immune system diseases, namely, Secukinumab and Ixekizumab (Anatomical Therapeutic Chemical [ATC] code L04AC10 and L04AC13, respectively). Given the role of IL-17A as a key cytokine in the pathogenesis of Chagas disease, the opportunities for drug repurposing becomes very important, as there are only two treatments available, Benznidazole and Nifurtimox, with high rates of adverse effects and treatment withdrawal (91, 175).

1.3.1. IL18

IL-18 is produced by a wide variety of cells, including dendritic cells, macrophages, keratinocytes, intestinal epithelial cells, and osteoblasts, with a key pathophysiological role in health and disease (176) This

interleukin also plays an important role in the regulation of IFN-γ production and development of Th1 response. Given the numerous roles of IL-18 in the immune response, several studies have highlighted its implication in the acute and chronic phase of Chagas disease (177-180). Considering that infectious diseases exert significant selective genetic pressure, it has been proposed two mechanisms to explain the implication of the genetic background in the pathogenesis of Chagas disease (137). First, pathogen resistance genes (PRG) would be involved in inhibits infection by directly reducing pathogen burden and secondly, disease tolerance genes (DTG) will operate to minimize tissue damaging effects of the pathogen (181-183). Consequently, polymorphisms in PRG and DTG will be associated with differential disease progression. One of the most relevant disease tolerance genes identified was related to directly or indirectly inhibit IFN-γ production or Th1 differentiation (137), and therefore, IL-18 could be implicated in this regulation.

Taking this into consideration, three *IL18* genetic variants were assessed in four Latin American populations, being the largest candidate gene study conducted in this gene and Chagas disease to date. These results evidenced the implication of the *IL18* rs360719 polymorphism in the susceptibility to the infection although, when comparing cardiomyopathy and asymptomatic patients, no significant associations were detected. The previous study in a Colombian cohort was the first to report an association for rs2043055, rs1946518 and rs360719 with *T. cruzi* infection, suggesting that this association was mainly driven by the polymorphism rs360719 (122). After the enlargement of this cohort, the association remained, showing consistent results in a well-powered cohort. Replication of these variants was performed in an Argentinian cohort and only rs2043055 showed a borderline genetic association but

in the opposite direction compared with the Colombian cohort. These differences could be due to the complex genetic structure of Latin American individuals, reflected by the recent admixture among Native American, European, and African source populations (152). Also, as might occur with *IL17A* polymorphisms, the lack of replication may occur if the assessed polymorphism is not the causal variant but is rather in LD. Although the rs360719 in the Argentinian cohort showed no association with *T. cruzi* infection, probably as consequence of an insufficient statistical power in this case, the meta-analysis showed that this variant presented a consistent effect among the two cohorts, indicating protection against *T. cruzi* infection. This variant is located in the promoter region of the *IL18* gene and its functional annotation revealed its location in histone marks in primary mononuclear cells and T helper naive cells from peripheral blood, also being described as transcription factor. In addition, it has been described that *IL18* rs360719 polymorphism leads to loss of the octamer 1 (OCT-1) transcription factor binding site. OCT-1 is known to be a ubiquitously expressed factor involved in the regulation of certain cytokines, like IL-18 (184).

Local expression of *IL18* in chronic Chagas cardiomyopathy heart tissue has been described and seems to be associated with mononuclear inflammatory infiltrates, cardiomyocyte destruction and fibrosis (171). Previous analyses of the *IL18* genetic variant rs2043055 in the Brazilian population showed nominal significant differences in the genotypic frequencies among moderate and severe chronic cardiac patients (109). In the present analyses, when compare chronic Chagas cardiomyopathy with asymptomatic patients the same variant showed significant association in the Colombian cohort. However, these results were not

validated in the Bolivian and Argentinian cohorts. These discrepancies in results could be due to the genetic heterogeneity among the study cohorts (152), as the impact of European colonization and slave trade from western Africa has been demonstrated to altered the genomes of Native Americans in multiple and dynamic ways (66, 185). Thus, previous results have showed that the prevalence of *T. cruzi* infection is strongly and independently associated with higher levels of African and Native American ancestry in a Brazilian population (186). This heterogeneity in ancestry proportions across geographic regions, and also within countries themselves, are challenging in association studies in order to find generalizable results across populations (187-189). All this, suggests that a fine-scale genomics perspective might represent a powerful tool to understand the role of genetics in this neglected disease diagnosis and prognosis.

In summary, these results validated the previous work on *IL18*, highlighting the role of rs360719 in the susceptibility to *T. cruzi* infection (122). Although, no evidence of association was found between the *IL18* genetic variants and chronic Chagas cardiomyopathy in the populations under study, the meta-analyses supposes a powerful approach to identify genetic variants that influence susceptibility of common diseases (190, 191). Thus, in the context of Chagas disease is necessary to contemplate the challenges of studying such an heterogeneous populations like Latin Americans with recent admixture, where fine-scale genomic assessments may be necessary (152). Therefore, further studies are needed to reach more conclusive results concerning the genetic variability in the *IL18* genomic region.

Chapter 2: Identification of novel susceptibility *loci* in chronic Chagas cardiomyopathy throughout a genome-wide association study

Complex diseases are influenced by the added contribution of several common genetic variants. In addition, the genetic architecture of these traits makes a challenge the assessment of its influence using candidate gene approaches. Thus, GWAS represent a great approach in the study of the genetic component of infectious diseases. For this, in order to better elucidate the genetic basis of Chagas disease and chronic Chagas cardiomyopathy, this chapter shows the results of the largest GWAS and meta-analysis of different Latin American populations carried out to the date in this disease. The meta-analysis revealed a novel genome-wide statistically significant association with chronic Chagas cardiomyopathy development nearby the *SAC3D1* gene that, in combination with the *in silico* functional relationships between the associated variant and the *SNX15*, *BAFT2*, and *FERMT3* genes, provides further insights into the pathogenesis of this neglected disease.

2.1. Materials and methods

2.1.1. Study populations, sample collection and ethical considerations

Samples from 3 different Latin American countries: Colombia, Bolivia, and Argentina were included in this study and meta-analyzed with data from the previous GWAS in a Brazilian population (192), comprising a total of 3,699 genomic DNA recruited samples. The description of sample origin is available elsewhere (193). Chagas disease cases were defined by their serological status using recombinant antigen and a commercial indirect hemagglutination test for ELISA assays. Participants were classified as seropositive and seronegative for *T. cruzi* antigens and comparisons were performed among them in order to assess susceptibility to the infection in the Colombian and Argentinian cohort. The Bolivian and Brazilian cohorts were composed only of seropositive individuals. In addition, all the seropositive patients recruited in this study were subjected to electrocardiograms, echocardiograms, and chest radiography to identify cardiac involvement. For the assessment of the chronic Chagas cardiomyopathy susceptibility, those patients with cardiac abnormalities were compared to asymptomatic individuals. Data included from the Brazilian cohort were described elsewhere (192). Sample size and demographic information are summarized in Table 2.1. The study was approved by the Ethics Committees from the Industrial University of Santander and Cardiovascular Foundation, Colombia (Act No. 15/2005), the Vall D'Hebron University Hospital, Barcelona, Spain (PR (AMI) 297/2016), and the National Hospital of Clinics, National University of Cordoba, Argentina (CIEIS HNC 118/2012 and 2/16/2017). The Ethics Committees for the Brazilian cohort was described elsewhere (192). Protocols used in

the study followed the principles of the Declaration of Helsinki, and all individuals included in the study signed written informed consents.

Table 2.1. Demographical characteristics and classification of data collections.

	Seropositive								Seronegative
	Colombia		Bolivia		Argentina		Brazil*		Colombia
	CCC	**ASY**	**CCC**	**ASY**	**CCC**	**ASY**	**CCC**	**ASY**	
Subjects, N	577	442	97	518	100	62	207	306	1,104
Sex (% females)	280 (48 %)	282 (64%)	57 (59%)	364 (70%)	65 (65%)	40 (65%)	81 (39%)	154 (53%)	672 (61%)
Age (Mean±SD)	67.4±14.3	57.2±15.6	50.7±9.4	46.4±9.5	58.2±9.4	47.5±14.0	-	-	49.5±16.7

The sample sizes refer to data passing the genotyping quality controls.

Abbreviations: CCC, chronic chagas Cardiomyopathy; ASY, asymptomatic.

* Data set from Deng et al, 2013

2.1.2. Genotyping, quality control and imputation

Genomic DNA from blood samples from the Colombia, Bolivian, and Argentinian cohorts were isolated using the *QIAamp Midi*DNA Kit following manufacturer's recommendations. The samples were genotyped using the Global Screening Array Platform (Illumina Inc., San Diego, CA, USA). Quality controls (QCs) of genotyped data and consistency of sex assignment were performed using Plink v.1.9 (125). Individuals and variants with missing genotype rate <5% were excluded. SNPs with different call rates between cases and controls (*P-value*<0.05), or deviating from HWE with a *P-value*≤1×10^{-06} in the control group were removed.

A principal component analysis (PCA) was performed using Plink1.9 to compare our data with the data from 1KGP. The genomic inflation factor (λ) was calculated to control the type I error rate with the package "gap" for R version 3.6.1. After QCs, genotyped data were imputed with the Michigan

Imputation Server using the admixed American population from 1KGP phase 3 as reference panel (54). After imputation, variants were filtered out by MAF<1% and the imputation quality metric Rsq<0.3.

2.1.3. Statistical analysis

For each case-control collection, association testing for imputed allele dosages was performed using a mixed model with EMMAX (194) implemented in EPACTS (https://genome.sph.umich.edu/wiki/EPACTS). Given the common environmental exposure to triatomine vectors of individuals living together in endemic areas, related individuals were maintained in the analysis. In this sense, genotyped data were used to calculate the pairwise identity by descent (IBD) proportion (PI_HAT), duplicated individuals were removed (PI_HAT>0.8) and related individuals were considered with a PI_HAT≥0.4. EMMAX calculates the kinship matrix to include it as a covariate in the statistical analysis, as well as sex and age. The summary level statistics of each cohort, including the Brazilian cohort, were meta-analyzed using METASOFT (126) under a fixed and random-effect models based on variants' heterogeneity (Cochran's Q test *P-value*). As most of imputed variants (>95%) did not show heterogeneity (Cochran's Q test *P-value*>0.05), those analyzed under a random-effect model were omitted from further analysis. Conservative suggestive associations were established at a *P-value*<1×10^{-06} and genome-wide significance at *P-value*<5×10^{-08} as usually employed for GWAS. The SNPnexus software (https://www.snp-nexus.org/v4/) was used to annotate the suggestive and significant signals, and LocusZoom (http://locuszoom.org/) for the regional association plots.

2.1.4. Functional annotation of associated variants

An *in silico* approach was used to investigate the potential biological consequences of the associated variant. The LDlink tool (195) was used to calculate proxies in high LD with the top signal (r^2>0.8) and different databases were queried. The Open Targets Genetic was used to assess their functional implication (https://genetics.opentargets. org/), and complemented with PheWeb version 1.1.17 (http://pheweb.sph.umich.edu/SAIGE-UKB/) to assess these variants in previous Phenome-Wide Association Studies (PheWAS). We used HaploReg v4.1 (196) to evaluate regulatory genomic regions and Capture HiC Plotter (197) to assess long-distance physical interactions. RegulomeDB (https://regulomedb.org/regulomesearch/) and GTEx (https://gtexportal.org/home/) were used to evaluate expression-quantitative trait *loci* (eQTLs) in relevant tissues.

2.2. Results

A total of 3,413 postQCs samples from 4 independent Latin American cohorts from Colombia, Bolivia, Argentina, and Brazil were included in this study. After QCs and imputation, 7,846,902 SNPs for the Colombian and 8,408,292 SNPs for the Argentinian cohorts were analyzed correcting by age, sex, and kinship in the case of susceptibility to the infection by *T. cruzi* analysis, while in the case of chronic Chagas cardiomyopathy development, a total of 8,218,190 SNPs were tested in the meta-analysis including the Colombia, Bolivian, Argentinian, and Brazilian populations.

2.2.1. Population structure

Given the ethnic admixture of Latin American populations, the first 2 principal components were used to evaluate the genetic heterogeneity existing in the samples included in this study and in comparison with the admixed American subpopulation from 1KGP. The results of these analyses showed that the heterogeneity present in our samples is consistent with the observed in the admixed American subpopulation of 1KGP. Despite this, the Argentinian and Bolivian sample collections seemed to be more homogenous, while a considerable genetic variability with the Brazilian cohort was detected, reinforcing the high degree of genetic admixture of this population (Figure 2.1). In the association analyses within each cohort, the effect of population structure was taken into account by using a mixed-model of association (194).

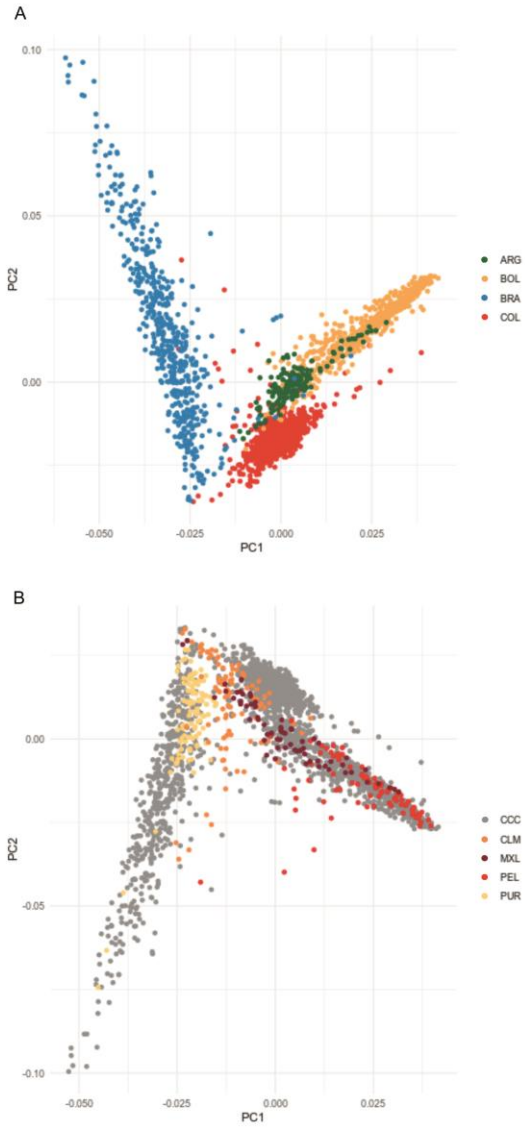

Figure 2.1. Principal component analysis of (A) study populations in comparison with (B) the Admixed American populations from the 1000 Genomes Project. Abbreviations: ARG Argentina; BOL Bolivia; BRAZ Brazil; COL Colombia. For 1000 Genomes Project populations: CLM Colombian; MXL Mexican; PEL Peruvian; PUR Puerto Rican.

2.2.2. Differential Susceptibility to T. cruzi infection

Individuals from the Colombian and Argentinian cohorts were classified and compared according to their serological status in seropositive and seronegative for *T. cruzi* infection. No statistically relevant results were identified in the association test in the Argentinian population, mainly due to the reduced sample size of this cohort. However, in the case of the Colombian cohort, we detected 7 suggestive associations (*P-value*<1×10^{-06}) contained in 4 different *loci* due to the LD (Table 2.2). The strongest association is an intronic variant of the *EBF2* gene (rs147475322; *P-value*=2.15×10^{-07}, OR=1.20, 95% CI 0.76–1.90). The other 3 SNPs are located in an intronic region of *CD247* gene (rs554994388) and map downstream and upstream of *SIK1* (rs229347) and *IL18* (rs4937075) genes, respectively (Table 2.2). Remarkably, a different SNP within the *IL18* gene was previously associated in a candidate gene assessment (193); however, this variant was nominally associated in our study and did not reach the suggestive level of significance.

Table 2.2. Suggestive* signals associated with susceptibility to *Trypanosoma cruzi* infection in the Colombian cohort

Chr	Locus	Bp	SNP	SNP status	Rsq	Effect allele	Allele Freq	*p-value*	OR (95% CI)	Function
1	*CD247*	167431884	rs554994388	Imputed	0.5	G	0.03	5.35E-07	1.33 (1.19-1.49)	Intronic
8	*EBF2*	25979367	rs147475322	Imputed	0.86	A	0.03	2.15E-07	1.2 (0.76-1.90)	Intronic
11	*IL18*	112127040	rs4937075	Imputed	0.77	C	0.77	8.34E-07	1.02 (0.86-1.21)	Intergenic
21	*SIK1*	43428190	rs229347	Imputed	0.88	A	0.6	8.93E-07	1.07 (1.04-1.10)	Intergenic

*Suggestive associations were considered with a *p-value* ≤1E-06
Associated *loci* were determined in order to the LD of contained variants and the most significant SNP of each *loci* were reported in the table.
The r-squared metric (Rsq) indicates the correlation between input genotypes and imputed dosages.

Abbreviations: Allele Freq allele frequency; bp base pair; chr chromosome; CI confidence interval; OR odds ratio; SNP info SNP information.

2.2.3. Differential susceptibility to the development of chronic Chagas cardiomyopathy

In order to identify novel genetic variants associated with differential chronic Chagas cardiomyopathy development, seropositive samples were stratified according to their cardiological status, and patients with cardiac abnormalities were compared to asymptomatic individuals in the association test. After QCs, genotype imputation, and filtering, a case-control association analysis from GWAS data was performed in the Colombian, Bolivian and Argentinian samples comprising a total of 2,309 individuals (981 chronic cardiac patients and 1,328 asymptomatic individuals). After correcting by age, sex, and kinship, deviations in genomic inflation factors were not observed. Next, summary statistics from every cohort, including the Brazilian data, were meta-analyzed by an inverse variance-weighted effect size and those shared by at least 2 data sets were considered. Additionally, we did not observe inflation in the meta-analysis (λ=1.03). As result, one statistically significant associated signal was identified at GWAS-level (*P-value*<5×10^{-08}) and 22 suggestive associations (*P-value*<1×10^{-06}) in 3 different autosomal *loci* (Table 2.3; Figure 2.2). The strongest association (rs2458298; *P-value*=3.27×10^{-08}, OR=0.90, 95% CI 0.87–0.94) is located in chromosome 11 in an intronic region of the *NAALADL1* gene. This signal is followed by several proxy variants in high or moderate LD (r^2>0.4) located nearby *NAALADL1*, *SAC3D1,* and *SNX15* genes (Figure 2.3). Regarding the rest of the suggestive signals, they are located in intergenic regions close to *CDH8* and *KLF4* genes (Table 2.3).

Table 2.3. Genome-wide significant association and suggestive* signals associated with chronic Chagas cardiomyopathy development in the meta-analysis.

Chr	Locus	Bp	SNP	SNP status	Rsq	Effect allele	Allele Freq	*p-value* Q	*p-value*	OR (95% CI)	Function
5	*CDH8*	18718887	rs10472156	Imputed	0.81	A	0.73	0.72	8.89E-07	1.1 (1.06-1.14)	Intergenic
9	*KLF4*	107479876	rs10759240	Genotyped	-	C	0.54	0.98	4.68E-07	0.92 (0.89-0.95)	Intergenic
11	*SAC3D1*	65047341	rs2458298	Imputed	0.95	G	0.26	0.98	3.27E-08	0.9 (0.87-0.94)	Intronic

*Suggestive associations were considered with a *p-value* ≤1E-06

Associated *loci* were determined in order to the LD of contained variants and the most significant SNP of each *loci* were reported in the table.

In boldface the genome-wide significant signal.

The r-squared metric (Rsq) indicates the correlation between input genotypes and imputed dosages.

Abbreviations: Allele Freq allele frequency; bp base pair; chr chromosome; CI confidence interval; OR odds ratio; *p-value* Q Cochran's Q test; rsq r-square; SNP info SNP information.

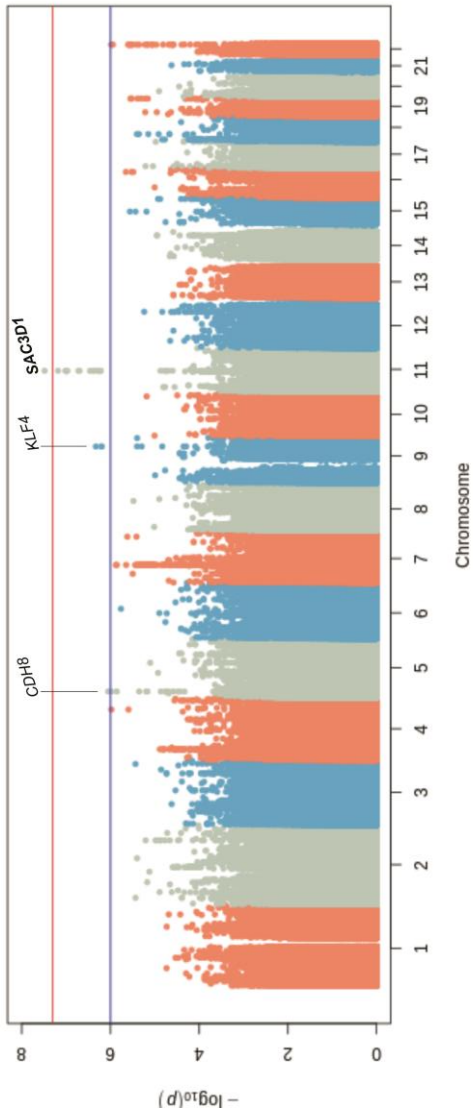

Figure 2.2. Manhattan plot of the association test in chronic Chagas cardiomyopathy meta-analysis results. Y and X axes refer the $-\log_{10}$ transformed P-values and positions in chromosomes, respectively. The red horizontal line refers the genome-wide association threshold, stablished at P-value<5×10^{-08}, and the blue line refers the suggestive threshold (P-value<1×10^{-06}).

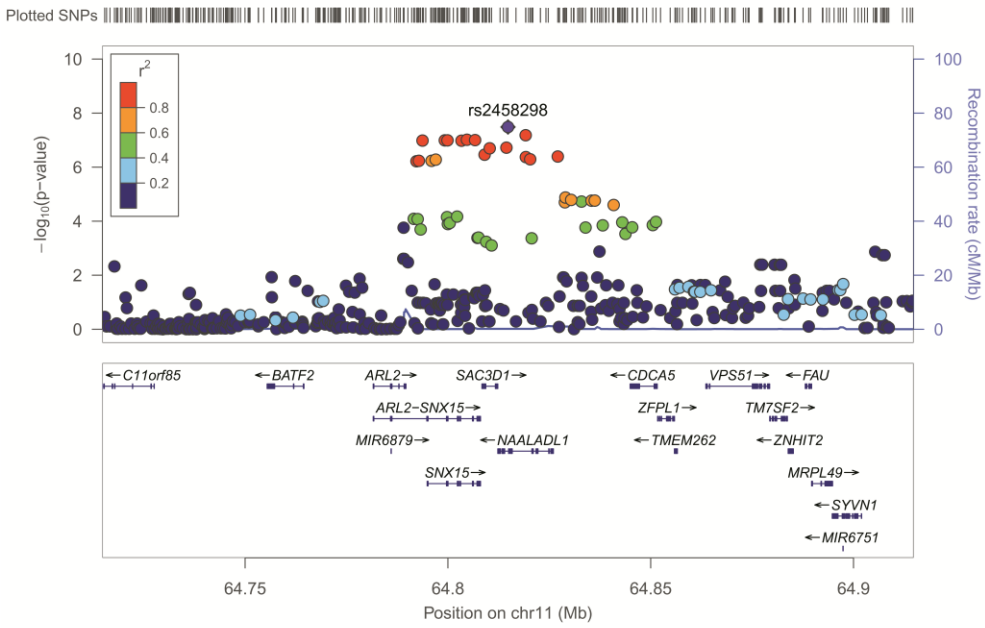

Figure 2.3. Regional association plot for the genome-wide significant signal. The –log10 transformed P-values of associated variants are plotted by position. The genomic position indicated correspond the top variant associated in this region. The rest of the variants are colored in function of their degree of LD with the associated signal. LD is based on pairwise r^2 values from the American population from the 1KGP. The blue line corresponds to the estimated recombination rates. Nearby genes to the associated and suggestive signals are shown in the bottom of the plot.

2.2.4. Functional analysis

To explore functional features of the genomic association obtained from the meta-analysis in chronic Chagas cardiomyopathy, several databases were queried. Regarding the top associated variant rs2458298, although is located in an intronic region of *NAALADL1*, it is closer to the transcription

start site (TSS) of *SAC3D1*, being this considered as its nearest gene. PheWAS information from Open Targets Genetics revealed a risk association of this SNP with cardiovascular traits such as hypertension (*P-value*PheWas=7.4×10^{-06}) and high blood pressure (*P-value*PheWas=5.8×10^{-06}) in the UK Biobank database, as well as with abdominal aortic aneurysm (*P-value*PheWas=1.9×10^{-03}) in a subset of patients. Significant eQTLs (i.e. *loci* that explain part of the variance in gene expression in specific tissues) were described identifying rs2458298 as correlated with *SNX15* expression in heart atrial appendage (*P-value* = 9.9 × 10-10) and heart left ventricle (*P-value*=1.3×10^{-11}). The same SNP was also correlated with the expression of *FERMT3* in a lymphoblastoid cell line. Using the Capture Hi-C Plotter platform, 42 interactions of this SNP with different gene regions across chromosome 11 were identified in the experiments done by Mifsud and colleagues (198) in a lymphoblastoid cell line (GM12878). One of those relevant promoter interactions is with the *BATF2* gene, which has been previously implicated in *T. cruzi* infection (199). All these results are summarized in Table 2.4.

Table 2.4. *In silico* functional analysis of the associated variant with chronic Chagas cardiomyopathy development

Chr	SNP	Nearest gene	eQTLs	C-HiC Genes	PheWAS
11	rs2458298	*SAC3D1* *NAALADL1*	*SNX15*[a] FERMT3[b]	*BATF2* in GM12878[b]	Hypertension, high blood pressure and abdominal aortic aneurysm

[a]eGENE in heart atrial appendage and heart left ventricle

[b]lymphoblastoid cell line

Queried data bases were Open Targets Genetics, GTEx, RegulomeDB, Capture HiC Plotter, and PheWEB

Abreviations: Chr chromosome; C-HiC capture Hi-C; eQTL expression quantitative trait loci; PheWAS phenome-wide association study.

2.3. Discussion

Several suggestive *loci* have been identified in association with differential susceptibility to *T. cruzi* infection in our study. Among these variants, highlight those located in intergenic regions near *IL18* and *CD247* given their implication in the immune response (200, 201). As described in previous chapters, *IL18* encodes a proinflammatory cytokine that is involved in the innate and adaptive response, and is critical for T-cell differentiation into IFN-γ producing Th1-type T cells and for IFN-γ production by NK cells (201). IFN-γ is a key player in the acute *T. cruzi* infection and a pathogen resistance gene, aiding the control of *T. cruzi* parasitism by stimulating surrounding cells to produce TNF-α and other inflammatory mediators leading to the generation of peroxynitrite (137). Additionally, IFN-γ gene expression has been reported to increase in chronic cardiac patients in several studies conducted in Latin American populations, indicating an important role of this molecule in the pathogenesis of the chronic cardiac form through the induction of inflammatory damage (202). Here, we confirm the association of this gene with the differential susceptibility to *T. cruzi* infection; although given the directed nature of candidate gene studies, different SNPs of the same locus were reported (122, 193). Regarding *CD247*, this gene encodes a subunit of the T-cell receptor CD3 complex that plays an important role in antigen recognition (200). In Chagas disease, previous works have confirmed lower expression of CD3 in isolated blood cells from individuals with high exposure to the parasite in Colombian populations (203). Finally, genetic variants in the *CD247* were also previously associated with autoimmunity (204-206) pointing to the existence of shared genes in autoimmune and infectious diseases as described elsewhere (207). Remarkably, a statistically significant association at genomic level was identified near the *SAC3D1* gene region (rs2458298) with the development of chronic Chagas cardiomyopathy.

SAC3D1, also known as *SHD1*, has been identified as a transcriptional regulator of *STAT5* (208), also associated with cardioprotection in humans (209). Additionally, the STAT-5 signaling by IL-2, IL-7, and IL-15 receptors has been shown to be perturbed in peripheral and heart-infiltrating T cells in chronic Chagas cardiomyopathy (210). An *in silico* functional assessment of the most associated variant and its proxies was performed in order to acknowledge the interactions and processes where they are involved.

Interestingly, physical contact of chromatin regions in this genetic location revealed interaction with *BATF2*. This gene encodes a transcription factor (TF) that has been related with Chagas disease, regulating the IL-23-Th17 pathway and suggesting an immunoregulatory function during *T. cruzi* infection in a mouse model (199). On the other hand, the top signal has been associated with several cardiovascular traits in PheWas data from UK Biobank. Additionally, this variant is an eQTL of *SNX15* in heart atrial appendage and heart left ventricle. This gene encodes a member of the sorting nexin protein family, which has been related with cardiovascular diseases (211). Additionally, the associated variant was also an eQTL for *FERMT3* (Fermitin Family Member 3) in lymphoblastoid cell lines. The *FERMT3* belongs to the Kindlins family and it encodes a protein that plays an important role in the regulation of thrombosis, as well as in maintaining the cytoskeleton of erythrocytes (212). This gene was previously associated with serum triglyceride levels, considered as a risk factor for coronary heart disease (213). Although the diagnosis of chronic Chagas cardiomyopathy involves the exclusion of other cardiomyopathies, the association with this gene may implicate common pathways among them. Other suggestive signals are those located in chromosome 9 near to *KLF4*, which encodes the Kruppel-like factor 4, a TF related with cardiac mitochondrial homeostasis in a mouse model (214), as well as with a high expression of the inducible

nitric oxide synthase and NO production in different cell types (215). Taking into account that higher production of NO has been reported in chronic Chagas cardiomyopathy severity (216), it would be possible a relationship among *KLF4* and chronic Chagas cardiomyopathy through the NO production.

This study has certain limitations. First, a cross-sectional assessment and the posterior seroconversion or development of cardiac events cannot be ruled out. In this case, the inclusion of cases in the control group will bias the results towards the null hypothesis of no association, weakening the risk of a false positive result. Second, the associated genes have been nominated as related with the disease through a functional assessment in publicly available non-endemic populations. Further functional analyses will guarantee the acceptance or rebuttal of these hypotheses. Finally, the small sample size in the overall study, but especially in the Chagas disease susceptibility assessment, bounds the statistical power for detecting lower frequency variants or with smaller effects. This is the reason why suggestive variants were highlighted, including their potential functional relevance. The main strength of our study is that we reveal a consistent and reproducible association with chronic Chagas cardiomyopathy in 4 Latin American populations. Genetic studies in admixed populations are challenging given their differential ancestry proportions (217), which is further complicated when taking also into account the genetic diversity of the infectious agent (13). This heterogeneity has been previously observed in a related parasitic disease such as Leishmania (83). Further studies evaluating the human–parasite genetic interactions may improve the knowledge of disease pathogenesis. Finally, the inclusion of underrepresented populations in genetic studies is a great opportunity to broaden the knowledge of complex diseases and deepen their molecular mechanisms (218), which is especially

important in Chagas disease, due to its significant socio-economic burden in both endemic and non-endemic countries (3).

We identified a novel locus near the gene region of *SAC3D1* associated with the chronic Chagas cardiomyopathy as well as several suggestive associations with the differential susceptibility to the infection by *T. cruzi* in several Latin American populations. These findings highlight the role of the host genetic component in Chagas disease, providing important novel leads to understand the pathogenesis of this neglected disease.

Chapter 3: Differential associations of genetic ancestries with Chagas disease susceptibility in the Colombian population using an admixture mapping approach

The complexity of the host genetic implication in infectious diseases, such as Chagas disease, makes possible the contribution population-specific polymorphisms to the disease risk, especially in recently admixed populations. This is the case of Latin American populations, which are a three-way admixture between Native American, European and African populations, being the Colombian the one with the most disparate admixture proportions. In order to assess genomic susceptibility regions where affected individuals share their genetic ancestry, this chapter shows the results of the first admixture mapping study carried out in Chagas disease for the Colombian population. A two-stage study was conducted and subjects were classified as seropositive and seronegative for *T. cruzi*, obtaining a protective association of the Native American ancestry with Chagas disease in the Major Histocompatibility Complex region. Additionally, a selective sweep scan in an AMR reference population from 1KGP together with an *in silico* functional analysis highlighted the Tripartite Motif family and the Human Leukocyte Antigen genes, with crucial role in the immune response against pathogens.

3. 1. Material and Methods

3.1.1. Ethical considerations

The protocols used in the study followed the Declaration of Helsinki principles and informed consent was obtained from all individuals included in the study design. The Industrial University of Santander and Cardiovascular Foundation (Colombia) Ethics Committee approved this study (Act No. 15/2005).

3.1.2. Study population and genotyping

All donors were recruited by the health care system from the Industrial University of Santander and Cardiovascular Foundation in the provinces of Guanentina, Comunera and Garcia Rovira, which are the provinces with the highest prevalence of Chagas disease in the Santander department in Colombia (219). We used a two-stage case–control design where individuals were classified as seropositive (cases) or seronegative (controls) for *T. cruzi* antigens according to an indirect hemagglutination commercial test (Chagatest, Wiener, Argentina) and enzymelinked immunosorbent assays (Test ELISA Chagas III, Grupo Bios, Chile; Chagas ELISA IgG + IgM, Vircell, Spain). Samples from stage 1 comprised 1,576 individuals (933 classified as cases and 643 as controls) as described elsewhere (220). Further sample recruitment composed stage 2, including 654 independent samples (122 cases and 532 controls) according to the same classification criteria. Genomic DNA isolation of blood samples was performed using the QIAamp Midi DNA Kit (QIAGEN, Germany) following manufacturer's recommendations. All samples were genotyped with the Global Screening

Array Platform (Illumina Inc., San Diego, CA, USA) as described elsewhere (221). As part of the QCs of genotyped data, individuals and variants with missing genotype rate >5%, SNPs with different call rates between cases and controls (*P-value*<0.05), and SNPs with large deviations from HWE in the control group (*P-value*≤1×10^{-6}) was removed. QCs were performed using PLINK v.1.9 (125).

3.1.3. Admixture mapping analysis

Reference data from the 1KGP Phase III (54) was used as representatives of the parental populations in the downstream stage 1 global and local ancestry assessments. Briefly, we used data from Native American (AMR, n=85), European (EUR, n=503) and African (AFR, n=504) reference populations. Regarding the AMR population, only Peruvian from Lima (PEL) were included because this population has been described to have the highest proportion of AMR ancestry after the Maya population (222). All EUR subpopulations were taken into account (Utah residents, Finnish, British, Iberian and Toscani populations). In the case of AFR, both the African Caribbean from Barbados, and individuals with African ancestry from southwest US were excluded as they are populations with recent admixture events (223). Using PLINK, we intersected the autosomal SNPs from stage 1 genotyped data (post QC 493,271 SNPs) with those from the 1KGP dataset, removing from the latter those variants with missing genotype rate >5% or with large deviations from HWE (*P-value*≤1×10^{-6}) in at least one population. After the intersection of datasets and QCs, a total of 471,342 SNPs were kept for further analysis.

Global ancestry assessment. Individual global ancestries were obtained with ADMIXTURE v1.3.0 (224), which calculates average ancestry

proportions across the genome. We estimated the best number of ancestral populations (K) using a 10-time cross validation with random seeds. The best fitting was obtained for K=3, i.e. assuming 3 ancestral populations distinguishing AMR, EUR and AFR ancestries.

Local ancestry assessment and association testing. Local ancestries were estimated with ELAI v1.0 (225) assuming a three-way admixture and 10 generations since the last admixture event as has been indicated by previous studies (66, 152). As QC steps, a correlation was calculated for individual global and averaged local ancestries across the genome using R v3.6.1. Principal components were calculated using PLINK and plotted along with those from the representative parental populations from 1KGP.

For the admixture mapping, the three local ancestries by separate were tested for case–control association using the EMMAX mixed model (194) implemented in EPACTS (https://genome.sph.umich.edu/wiki/EPACTS). As related individuals were considered, EMMAX calculates the kinship matrix to include it as a covariate in the model, in addition to age and sex. We controlled for the type I error rate by calculating the λ using an in-house script for R v3.6.1. Significance was adjusted by the number of ancestry blocks across the genome and the number of generations since the admixture using the R package STEAM (66). Based on that, the significance threshold for the admixture mapping study was established at *P-value*<3.07×10^{-06}, similar to estimates that have been declared in independent studies in Latin American populations (66).

3.1.4. Fine mapping

A fine mapping assessment was performed to elucidate the most likely genetic variant(s) underlying the admixture mapping study results combining data from the two stages. Briefly, genotypic data from stage 2 were subjected to the same QCs that were used in stage 1. After this, imputation of both stages was performed with the Michigan Imputation Server using the admixed American population from 1KGP phase 3 as reference panel. Imputed variants were filtered by their MAF and the imputation quality metric Rsq. Those variants satisfying both a MAF>1% and Rsq>0.3 were kept for the study. Association testing of imputed allele dosages was performed in both stages by separate using the EMMAX mixed model implemented in EPACTS, including age, sex and the kinship matrix as covariates. Summary-level statistics of each stage were meta-analyzed using METASOFT v2.0 (126). A random or fixed-effect size meta-analysis was selected for each variant based on the results for the Cochran's Q-test of heterogeneity. Significant and suggestive thresholds for the assessed region were established at *P-value*<3.23×10^{-04} and *P-value*<6.46×10^{-03}, respectively, according to the estimates of GEC software (226) based on the LD structure from the stage 1 samples. In order to confirm the association of the specific allele with the local AMR ancestry, samples were stratified according to their ancestry punctuations and allele frequencies were recalculated in cases and controls separately. The statistical power of the fine mapping analysis was estimated using the Power Calculator for Two Stage Association Studies software CaTS (127).

3.1.5. Selective sweep analysis

Given the high rate of adaptive signals in the genome triggered by parasites, and that admixture serves as a mechanism driving adaptive evolution in humans (68, 227), we used iSAFE v1.0.4 (228) to provide evidence of a selective sweep embedded in the admixture mapping region and to pinpoint the most likely favored variant. iSAFE exploits the evolutionary contributions hidden in the flanking regions surrounding the region under selection to provide a ranking of variants (iSAFE-score) based on their contribution to the overall signal of selection. For the analysis, we used 1KGP data from unrelated subjects from PEL population (n=77) and from a random selection of 10% Yoruba individuals drawn from the reference (n=91), which represents a non-target or outgroup population. iSAFE was executed enabling the IgnoreGaps flag and the default MaxFreq value (0.95). Ancestral fasta sequences for Homo sapiens (GRCh37) were downloaded from ENSEMBL release 75 (http://ftp.ensembl.org/pub/release-75/fasta/ancestral_alleles/).

3.1.6. In silico functional analysis

In silico functional analyses were performed to assess the biological consequences of the leading associated variants in the fine mapping and of those variants prioritized in the selective sweep analysis. The Open Targets Genetics portal (https://genetics.opentargets.org/) was used for functional annotation, to assess trait associations based on PheWAS and to retrieve the evidence of eQTLs in relevant tissues based on eQTLGen database (https://www.eqtlgen.org/). Additionally, long-distance physical interactions and regulatory genomic regions were considered using Capture HiC Plotter (197) and HaploReg v4.1 (196).

3.2. **Results**

This study has a two-stage case–control design and samples were classified as seropositive (cases) and seronegative (controls) for parasite antigens. Demographic characteristics are summarized in Table 3.1. Cases and controls from the stage 1 samples were matched in terms of global ancestries (Figure 3.1). Local ancestry blocks were estimated for the 471,342 positions corresponding to those of genotyped SNPs in stage 1. Local ancestry punctuations per individual were averaged and compared with their global ancestries showing a high correlation among them (r=0.89–0.96) and indicating consistent global and local ancestry proportions in the population under study. Cases and controls were also matched in terms of local ancestries, and supervised and unsupervised analysis provided consistent results, reinforcing the selection of the representatives of the parental populations in the ancestry assessment.

Table 3.1. Demographical characteristics and sample size of the Colombian collections.

	Stage 1		Stage 2	
	Seropositive	Seronegative	Seropositive	Seronegative
Pre QC Sample size	998	659	122	532
Post QC Sample size	913	592	122	512
Sex (% females)	503 (55%)	357 (60%)	65 (53%)	315 (62%)
Age (Mean±SD)	62.7±16.2	49.0±17.6	64.7±11.2	50.0±15.6

Genotyping quality controls (QC).

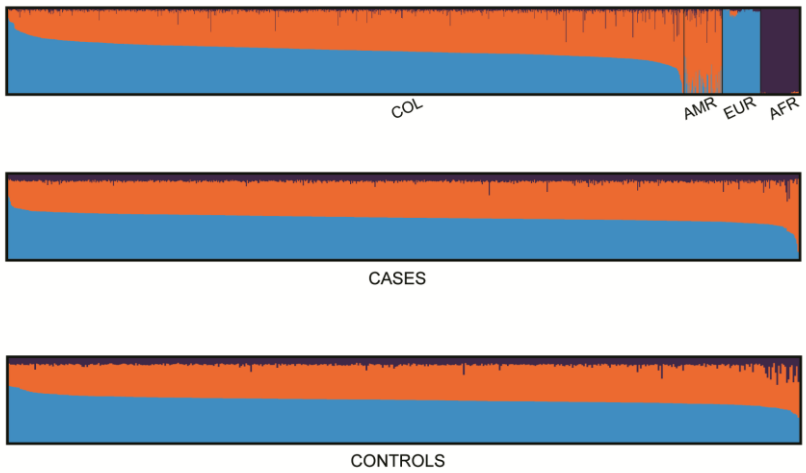

Figure 3.1. Admixture estimations (K=3) for individuals included in Stage 1 (Chagas disease cases and controls) and for AMR, EUR and AFR reference populations from 1KGP.

Admixture mapping results based on local ancestry estimations were not inflated because of the presence of population stratification for any of the ancestries after λ correction (λ_{AMR}=1.00, λ_{EUR}=1.00, λ_{AFR}=1.00). These results revealed genome-wide significant associations of AMR and EUR ancestries with Chagas disease in the positions of the chromosome 6 region 30,079,993-30,332,160 according to build hg19 (Figure 3.2). We observed a lack of association of AFR ancestry with the serological status. These results revealed a differential susceptibility to the infection associated with two of the ancestries in this particular region, where AMR was in higher proportion among seronegative individuals overall, therefore associated with a protective effect. The associated region is located in the MHC locus where the leading signal corresponds to rs115833233 (AMR ancestry OR=0.74, 95% CI=0.66–0.83, *P-value*=4.53×10^{-08}), which is located in the untranslated region (UTR) of exon 5 in the Tripartite Motif Containing 40 (*TRIM40*) gene. The association of AMR ancestry was unrelated with the genotypes in that position because conditioning by the allele dosage of the rs115833233 variant did not change the results. Thus, the admixture mapping peak was not explained by the genetic variation of that SNP (Table 3.2).

Table 3.2. Joint SNP-ancestry analysis in the discovery stage.

Factor	OR (95% CI)	*p-value*
AMR ancestry	0.74 (0.66-0.83)	4.53E-08
Allele dosage rs115833233	1.00 (0.89-1.13)	0.99
AMR ancestry (conditioned on rs115833233)	0.73 (0.65-0.82)	2.15E-08

Abbreviations: AMR Native American ancestry; CI confidence interval; OR odds ratio

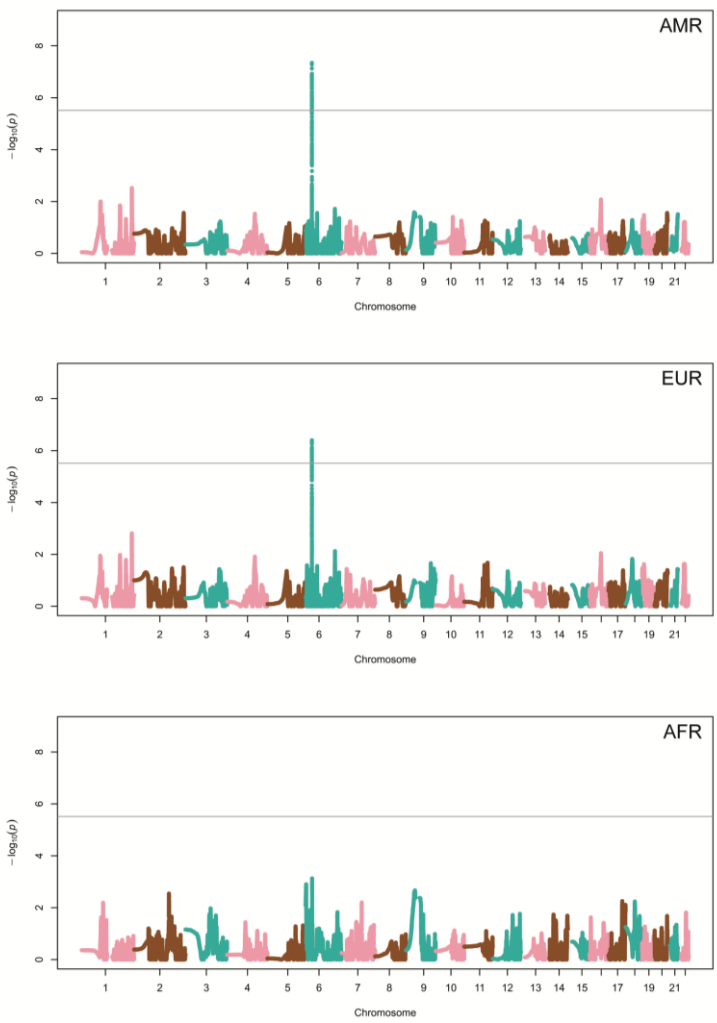

Figure 3.2. Manhattan plots of the admixture mapping results based on local ancestry estimates of Native American (AMR; up), European (EUR; middle) and African (AFR bottom). Y and X axes refer to the –log10 transformed P-values and hg19 positions in chromosomes, respectively. The horizontal line indicates the significance threshold (*P-value*=3.07 × 10^{-06}).

To identify potential variants explaining the result, imputed genotype data from the significant admixture mapping region were assessed in the stage 1 samples and in independent samples from stage 2. This analysis identified 23 variants in high LD (r^2>0.8) with nominal significance and consistent direction of effects in the two stages (Table 3.3). The leading variant was rs2032134, located intergenic to the Ribonuclease P/MRP Subunit P21 (*RPP21*) and *TRIM39* genes (OR for the C allele =0.93, 95% CI=0.90–0.97, *P-value*=3.54×10^{-04}) (Figure 3.3). We confirmed this association was dependent on the AMR ancestry because adjusting the models by the ancestry score resulted not significant (OR for the C allele=0.99, 95% CI=0.95–1.03, *P-value*=0.766). In addition, we confirmed that the C allele was associated with local AMR ancestry as stratifying cases and controls by this ancestry proportion, the allele frequency was increased among carriers of local AMR ancestry in cases and controls in comparison with individuals bearing other ancestries in that position. Functional analysis showed that rs2032134 is an eQTL for different HLA members in different tissues, most significantly for the *HLA-C* in whole blood (*P-value*=5.1×10^{-35}) according to eQTLGen. Evidence of long-distance interactions between rs2032134 and another member of the Tripartite Motif (TRIM) family, *TRIM31*, was observed in macrophages and neutrophils (229), as well as in lymphoblastoid cell lines (198). Further functional assessments for rs2032134 and its best LD-proxies are summarized in Table 3.4.

Table 3.3. Association testing results and allele frequencies on imputed data for the admixture mapping associated region.

		Stage 1			Stage 2			Meta-analysis	
SNP_ID(*EA)	CHR:BP	EA frequency (cases/controls)	OR (95%CI)	*p-value*	EA frequency (cases/controls	OR (95%CI)	*p-value*	OR (95%CI)	*p-value*
rs2032134*T	6:30360509	0.19/0.14	1.05 (1.01-1.10)	2.57E-02	0.25/0.16	1.09 (1.03-1.15)	3.67E-03	1.07 (1.03-1.11)	3.54E-04

Abbreviations: EA effect allele; BP base pair; CHR chromosome; CI confidence interval; OR odds ratio; SNP single nucleotide polymorphism.

Table 3.4. *In silico* functional assessment of the fine mapping and selective sweep analyses top variants.

Chr	SNP	Function	Nearest gene	eQTLs[a]	C-HiC genes	PheWAS
6	rs2032134	Intergenic	*RPP21* / *TRIM39*	*HLA-C*, *HCG18*, *HLA-E*, *HLA-G*, *TRIM10*	*TRIM31* in GM12878[b], macrophages[c] and neutrophils[c]	-
6	rs9261440	Intergenic	*TRIM31* / *TRIM40*	*HLA-F*, *HLA-G*, *HLA-A*, *HLA-L*, *TRIM26*, *TRIM10*	*TRIM10* in GM12878[b]	Hematological measurement, eosinphil count

a eGENES in whole blood

b lymphoblastoid cell line from Mifsud et al.

c Macrophages and neutrophil from Javierre et al.

Queried data bases were Open Targets Genetics, eQTLGen, HaploReg and Capture HiC Plotter

Abbreviations: Chr chromosome; C-HiC capture Hi-C; eQTL expression quantitative trait *loci* ; PheWAS phenome-wide association study.

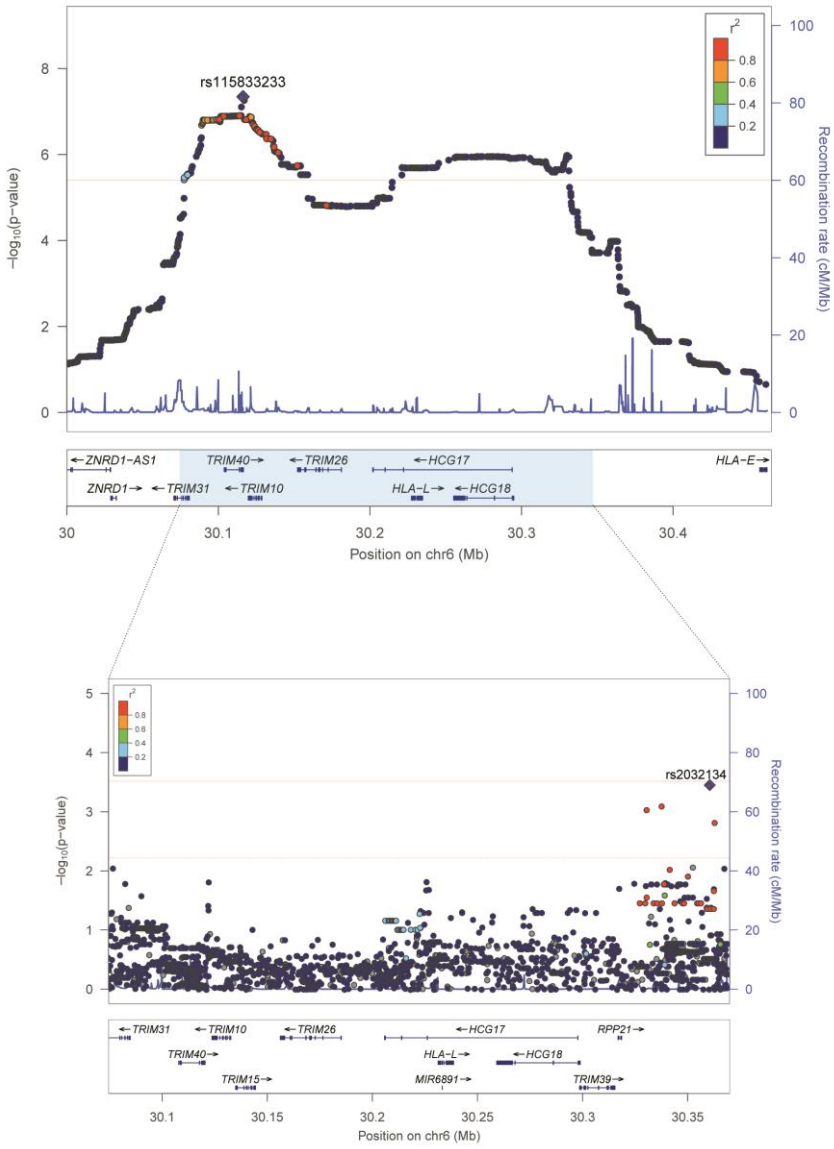

Figure 3.3. Regional plots of the results from admixture (top) and fine mapping (bottom) analyses, showing the $-\log_{10}$ transformed P-value in the y-axis and the hg19 genomic position in the x-axis. Top: Region centered on the significant variants

for the AMR ancestry association. The horizontal line indicates the significance threshold (*P-value*=3.07×10^{-6}). Estimated recombination rates (light blue line) are plotted on the right y-axis. Bottom: Meta-analysis results from stages 1 and 2 within the admixture mapping significant region (chr6:30,079,993-30,332,160) and its proxies (r^2>0.8) indicating the associated variant with the lowest significance in the region (rs2032134). The results for the remaining single nucleotide polymorphisms (SNPs) are color coded to reflect their degree of LD with the indicated SNP based on pairwise r^2 values in AMR. The horizontal lines indicate the significant (solid line, *P-value*=3.23×10^{-04}) and suggestive thresholds (broken line, *P-value*=6.46×10^{-03}).

In order to assess adaptive signals within the significant admixture mapping region, we performed a selective sweep analysis in this region using a reference population. From this analysis, we found that the iSAFE scores were high; with the top ranking variants corresponding to positions 30,087–30,102 Mb (iSAFE score≥0.331). The top 25 variants ranked by the iSAFE score mapped within (n=5; only rs28400887 was coding but synonymous) or nearby (n=20) the *TRIM31* gene, with the furthest SNP located at 21.1 Kb from the gene (Figure 3.4). Given the excellent performance of iSAFE in prioritizing the most likely favored variant in 94% of the times (228), we performed a functional assessment of the variant with top scoring (rs9261440; iSAFE score=0.340) assuming it to be the variant driving the selective sweep. According to Open Targets Genetics portal, rs9261440 is an eQTL for some HLA class I genes in whole blood, where the most significant were *HLA-F* (*P-value*=5.4×10^{-143}), HLA-G (*P-value*=2.4×10^{-280}) and *HLA-A* (*P-value*=2.4×10^{-156}), and also for the TRIM family member *TRIM26* (*P-value*=7.9×10^{-79}). Additionally, HaploReg indicates that this variant is an eQTL of HLA-L and *TRIM10* in whole blood (230). In agreement with this, PheWAS data from Open Targets Genetics indicates that this variant associates with hematological measurements,

particularly with the eosinophil count in the UK Biobank (*P-value*=7.6×10^{-43}).

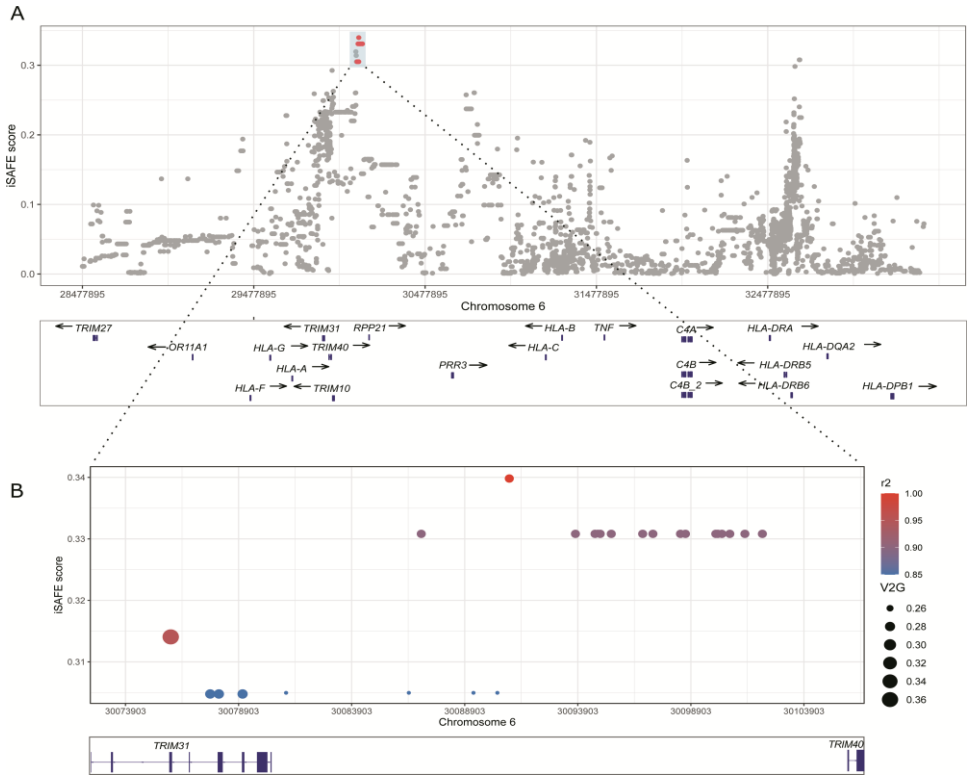

Figure 3.4. Selective sweep scan results for Native Americans (AMR) from the 1000 Genomes Project. Y and X axes represent the iSAFE scores and hg19 genomic positions, respectively. (A) iSAFE scores within the MHC region (chr6:28,477,895-33,389,603). Top ranking variants included in the significant admixture mapping region are highlighted in red (chr6:30,079,993-30,332,160). (B) Zoom in of the admixture mapping significant region, where the top 25 variants according to their iSAFE score are represented. Genetic variants are depicted in colored dots to ref lect its LD with the variant with the highest iSAFE score based on pairwise r^2 values in AMR population. Their dot size correlates to their composite variant to genes score (V2G) from Open Targets Genetics (https://genetics.opentargets.org/).

3.3. Discussion

In this chapter we describe the admixture mapping strategy to leverage varying local genomic ancestries in Colombians to identify *loci* associated with differential susceptibility to Chagas disease. Significant associations of AMR and EUR ancestries within the MHC locus with the development of the infection were found, pointing out to the role of the immune response on the disease risk. Additionally, fine mapping assessments and a selective sweep scan of this region prioritized variants with potential functional implications in the disease and highlighting how powerful this strategy is for identifying regions that had been previously overlooked in other genomic studies.

Several studies in bioarchaeological material confirmed that the American trypanosomiasis already existed long before European settlement when ancestral populations domesticated plants and animals in the process of sedentarization. This provided the vector with food availability and a more rapid domiciliation (3, 5). Host–parasite co-evolution is considered one of the most important generators of biological diversity in the genome, as is the case of selective sweeps at *loci* with functional role in their interaction (227). It occurs when parasites trigger host adaptations, which will lead the parasite to adapt again to this new environment in their hosts (227). This coevolution has been described for *T. cruzi* (39, 231) and, based on the observed protective association of the AMR ancestry, one can speculate that the long-term exposure of ancestral populations with larger proportions of AMR ancestry may have provided a more efficient immune response to the parasite, as previously hypothesized, with an adequate defense against pathogens that are endemic to the New World (68). This response could be responsible for parasite clearance, precluding the establishment of a serological response. This ancestry-specific effective

response against parasites and other infectious agents has been previously hypothesized to explain the enrichment of African alleles within the *HLA-B* locus in a Colombian population (67). Another study correlated the EUR and AFR ancestries with differential immune response to bacterial infections and hypothesized that this could be related to the differential pathogen exposure of each population and different selective pressures after the human migrations out of Africa (232). Local adaptations have shaped human genetic variation together with drift and migrations, and admixed populations are likely to have a larger number of genetic variants that have functional effects (68, 233). Therefore, it is particularly striking that the selective sweep scan within the significant admixture mapping region revealed variants associated with eosinophil counts because eosinophils have cytotoxic functions to fight parasitic infections, and along with macrophages, monocytes and neutrophils, are the innate immune cells responsible for the control of the initial infection by *T. cruzi* (234, 235). Interestingly, blood cell traits have been reported to differ among ancestries and are subjected to different selective pressures (236).

Several genes of the TRIM family mapped within the significant admixture mapping region, and the top variant was located in the *TRIM40* gene (Figure 3.3). More importantly, the variants prioritized both by the fine mapping and the selective sweep scan are also eQTLs of TRIM genes in whole blood. The TRIM family is E3 ubiquitin kinases that play an important role in immune signaling pathways, and the expansion of this multigene family suggests their key regulatory role during the immune response against pathogens (211, 237). When the pathogens are recognized by the immune system through the pattern-recognition receptors, several immune responses are initiated, including the production of interferons (IFNs), leading to the expression of TRIM proteins, among others (237). The upregulation of TRIM genes in response to *IFN-γ* has been reported in

human monocytes and macrophages (238). The expression of *IFN-γ* is induced by IL-18, which is a pro-inflammatory cytokine mainly produced by macrophages (39). Remarkably, genetic variants of the gene encoding *IL-18* have been associated with *T. cruzi* infection in previous candidate gene studies and showed suggestive association in a recently published GWAS (122, 193, 221).

The fact that the top variant from the fine mapping analysis mapped in an intergenic region, made us speculate on its potential relation with nearby genes through an in silico functional analysis. This strategy suggested the significant correlation of this variant with the *HLA-C* and *HLA-G* expression in whole blood as eQTLs. HLA genes are well-known for their role in the modulation of the immune response during *T. cruzi* infection (36). A previous study associated the HLA-C∗03 allele with the susceptibility to the chronic cardiac form of the disease in a Venezuelan population (239). Regarding the *HLA-G*, several alleles from the HLA-G*03 UTR region were tested in a Brazilian population, reporting evidence of association with different clinical forms of the disease (93). Additionally, lower gene expression and plasma concentrations of this gene have been described in the chronic phase of *T. cruzi* and Plasmodium falciparum infections (240). Further long-distance chromosome interaction analysis of this variant revealed a significant interaction with the promoter of *TRIM31* in macrophages, neutrophils and the lymphoblastoid cell line GM12878. This gene is a member of the TRIM family that acts as a regulator of the NLR pyrin domain-containing 3 (*NLRP3*) expression (241). The *NLRP3* is an inflammasome component that is well-known to be activated by molecular patterns associated with pathogens. The activation of the inflammasome is crucial for the control of intracellular protozoan parasitic infections, like *T. cruzi* and the production of nitric oxide during its acute infection in mouse models (216, 242). Interestingly, *TRIM31* has been also prioritized by the

selective sweep scan as a locus under putative positive selection in the AMR population.

There are some major limitations in this study that need to be taken into consideration. One of those limitations relates to the identification of the causal agent of the selective signal. The local adaptive signal maps within the MHC locus, which is a well-recognized target of selection. However, the causal factor underlying the selective sweep has been identified only in few instances in humans, such as skin pigmentation (243), lactase persistence (244) and adaptation to altitude (245). Therefore, we cannot infer nor guarantee that *T. cruzi* is the causal factor driving this putative adaptive signal. Another limitation is the limited statistical power, especially in the fine mapping, where we only had 80% power to detect variants with an allele frequency >20% and minimum effect allele of 1.4, which is fairly larger than expected for common variants in complex traits (246). Larger sample sizes would improve the statistical power of the study. Another limitation refers to the analyzed genetic variants, as only common SNPs were assessed in the fine mapping, and structural or less frequent genetic variants underlying the admixture mapping signal of Chagas susceptibility remain unexplored. Further insights concerning these types of genetic variants may be assessed by next-generation sequencing approaches. Moreover, despite the advantages of admixture mapping studies for identifying susceptibility loci, we are not in the position to warrantee if the results are generalizable to other populations with AMR ancestry, considering as well the variability of the infectious agent throughout the American continent. Finally, the scarce representation of AMR populations in reference databases reduces the precision to provide the exact functional implication of the associated variants, given the differences in LD structure among populations as well as other potential phenotypic features in different cell types.

This chapter encompasses the results of the first admixture mapping analysis carried out in Chagas disease. This assessment allowed us to associate the AMR local ancestry at the MHC locus with protection from Chagas disease, highlighting the role of the immune response during the acute phase of the infection.

Chapter 4: Functional assessment of the GWAS associated *loci* through its influence on methylation levels

GWAS associated and suggestive variants use to map in non-coding regions of the genome, which exert their effect modulating gene expression. This modulation might be driven by the methylation of the DNA through methylation quantitative trait *loci* (mQTL), which are correlations among SNPs and DNA methylation levels in the context of a specific trait. Thus, in order to functionally characterize GWAS associated variants that map in regulatory regions, these multi-omics integration analyses are performed. The aim of this study was to evaluate the functional mechanism underlying the previously reported region associated with chronic Chagas cardiomyopathy by a mQTL analysis in a subset of individuals from the Bolivian population. As result, we identified mQTLs in the *CCDC88B*, *PLAAT3* and *POLA2* genes, which have been associated with cardiovascular diseases in previous studies. Interestingly, the *CCDC88B* was previously identified as differentially methylated and expressed in heart biopsies of chronic Chagas cardiomyopathy patients. These results, in addition to suggest novel genes that could play a role in the chronic Chagas cardiomyopathy pathogenesis, evidence the functional relevance of the previously associated *loci*.

4.1. Material and methods

4.1.1. Study samples and ethical considerations

All donors were recruited by the health care system from the Vall D'Hebron University Hospital, Barcelona, Spain. The study population is composed by a subset of patients from Bolivia from the previous GWAS (221). A total of 57 seropositive individuals for *T. cruzi* antigens were classified into chronic Chagas cardiomyopathy patients (cases, n=37) and asymptomatic (controls, n=20) according to the presence of cardiac abnormalities. Patients were subjected to electrocardiogram and echocardiogram tests to determinate cardiac abnormalities, while additional clinical information or complementary tests were retrieved from the medical history. Case-control sample size and demographic information are summarized in Table 4.1.

Table 4.1. Demographical characteristics and sample size.

	CCC	ASY
Subjects, N	37	20
Sex (% females)	18 (48.65%)	16 (80%)
Age (Mean±SD)	52.8±9.9	45.5±12.9

The sample sizes refer to data passing the genotyping quality controls. Abbreviations: CCC, chronic chagas Cardiomyopathy; ASY, asymptomatic.

This study was approved by the Ethics Committee from the Vall D'Hebron University Hospital, Barcelona, Spain (PR (AMI) 297/2016). The Ethics Committees for the GWAS data was described elsewhere (221). Protocols

used in the study followed the principles of the Declaration of Helsinki and all individuals included in the study signed written informed consents.

4.1.2. Data preparation

SNP genotyping and imputation

As extensively described elsewhere (221), blood DNA was isolated and genotyped using the Global Screening Array Platform (Illumina Inc., San Diego, CA, USA). After passing their corresponding QCs using the software PLINK v.1.9 (125), genotyped data was imputed with the Michigan Imputation Server (247).

After imputation, genomic information from the 2 Mb flanking region centered in the variant associated with the disease (rs2458298, chr11:63,814,813-65,814,813, according to build hg19) was extracted from a subset of individuals from the Bolivian cohort, and transformed to PLINK format. Imputed SNPs were filtered by their imputation quality metric Rsq and MAF, keeping for the analysis those that satisfied both MAF>5% and Rsq>0.3.

DNA methylation data preprocessing

Genomic DNA from blood samples from the 57 Bolivian individuals was isolated using the QIAamp MidiDNA Kit (QIAGEN, Germany) following manufacturer's recommendations. From this, methylation patterns were determined using the Infinium MethylationEPIC Bead Chip array (Illumina, Inc., San Diego, CA, USA). A total of 850,000 DNAm sites were assessed and passed to downstream data processing and normalization using the R/Bioconductor package *minfi* (248). Probes positioned in SNP positions and those with a detection *p-value*<0.01 were removed. The clustering of methylated and unmethylated probes did not show significant differences in

the samples. After this, the preprocessQuantile method for normalization was performed (249) and the *M* values matrix was obtained by the $\log_2$-transformation of the DNA methylation ratio. This matrix was utilized to compare sample groups (cardiomyopathic and asymptomatic) by an eBayes-moderated paired t test using the limma package (250). As for the genomic data, methylation data was extracted for the same 2 Mb region centered in the variant associated with the disease and *p-value* of <0.05 was considered statistically significant.

4.1.3. Methylation quantitative trait loci statistical analysis.

In order to link genetic variants to variations in DNAm, preprocessed methylation and imputed genotyped data from the significant region were integrated for the 57 individuals included in this study. The analysis of *cis*-mQTLs for cases and controls was performed using the R/Bioconductor package *MatrixEQTL* version 2.3 (251). For the correlation analyses, sex and age were used as covariates and those SNP-DNAm site pairs with a maximum distance of 1 Mb were tested. In order to assess the functionality of the locus associated with the cardiac form of Chagas disease, the integration analysis was carried out in the region of 2 Mb centered in the associated GWAS variant (rs2458298, chr11:63,814,813-65,814,813 according to build hg19). Those interactions with a FDR of <0.05 were considered as significant.

4.1.4. In silico functional analyses

An *in silico* functional analysis was performed to assess the biological consequences of the genes related with the leading significant interactions. For this, genes within the DNAm sites were annotated using the R/Bioconductor package IlluminaHumanMethylationEPICmanifest (https://bioconductor.org/packages/release/data/annotation/html/IlluminaHumanMethylationEPICmanifest.html).

In order to prioritize *cis*-mQTLs linked to the variant associated with the disease, the LD link tool (195) was used to calculate SNPs in LD (r^2>0.4) with it in the admixed American population from The 1KGP. The Open Targets Genetics web tool (https://genetics.opentargets.org/) was used to evaluate the biological and disease implications of the genes associated with the DNAm sites from the prioritized *cis*-mQTLs.

4.2. Results

4.2.1. Identification of cis-mQTLs in the previously associated chronic Chagas cardiomyopathy locus

Genomic and methylation data from a total of 57 Chagas disease patients were integrated to determine *cis*-mQTLs. We limited our analysis to the genomic region centered in the variant previously associated with the chronic cardiac form of the disease (221). A total of 2,611 SNPs were tested against 2,647 DNAm sites in 37 chronic Chagas cardiomyopathy patients and 20 asymptomatic individuals after passing their corresponding QC criteria. We identified 6,958 significant *cis*-mQTLs (FDR<0.05) at maximal 1 Mb distance of the GWAS leading variant (Figure 4.1). These *cis*-mQTLs were composed by 2,143 unique SNPs and 152 unique DNAm sites. On average, the distance between the SNPs and DNAm sites that form the *cis*-mQTLs was ~114 kb.

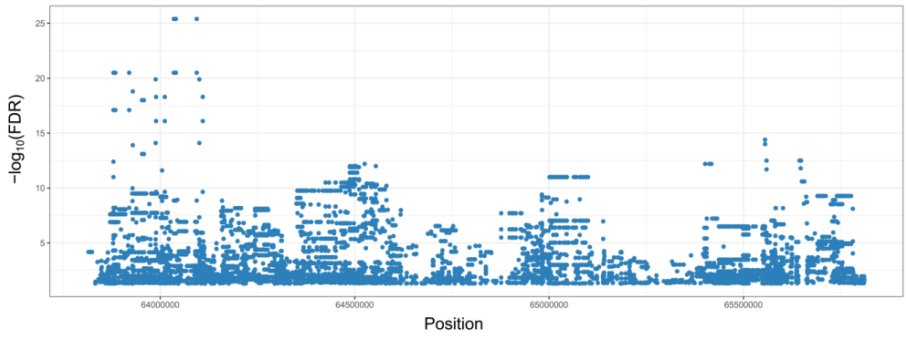

Figure 4.1. Regional scatter plot of the *cis*-mQTL results for the selected genomic region under analysis (chr11:63814813-65814813). Representation of the 6,958 *cis*-mQTLs identified in the genomic region previously associated to the disease. The x-axis corresponds to the genomic position of the SNP from each mQTL interaction; the y-axis refers to the transformed False Discovery Rate (FDR) of each signal.

First, we evaluated if the chronic Chagas cardiomyopathy associated variant and those in LD acted as *cis*-mQTLs. We identified six SNPs with moderate LD (mean r^2=0.55) with the leading variant as *cis*-mQTLs of a DNAm site located between 200-1,500 bp of distance to the TSS (TSS1500) of the SAC3 Domain Containing 1 (*SAC3D1*) gene (Table 4.2). The risk alleles of these variants were associated with higher methylation levels. However, these changes were not significant when comparing the methylation status in asymptomatic and cardiomyopathic patients. Nevertheless we observe a highly significant positive correlation of GWAS and *cis*-mQTL *p-values* (Figure 4.2) suggesting an association of both methylation and genetic signals. The most significant *cis*-mQTL in the region, which mapped in the gene body of the Phospholipase C Beta 3 (*PLCB3*), also showed no significant methylation changes. Therefore, although we observe a significant correlation of the allele changes with the methylation levels, we cannot reassure that this is directly related to the disease as this alteration in

methylation levels occurs in a similar manner in asymptomatic and CCC patients.

Table 4.2. ***cis*** **-mQTLs with SNPs in moderate LD (r^2>0.4) with the genetic variant associated to the chronic chagas cardiomyopathy.**

SNP ID	SNP position	DNAm site Id[a]	DNAm position	DNAm site function[b]	DNAm site gene	FDR	Beta	R^{2c}	Distance (bp)[d]
rs11231929	11:64810771	cg22690720	11:64808083	TSS1500	*SAC3D1*	1.85E-05	0.26	0.49	2,688
rs496427	11:64838173					6.82E-05	0.26	0.46	30,090
rs947799	11:64851305					8.06E-05	0.26	0.45	43,222
rs475089	11:64832939					1.46E-04	0.26	0.44	24,856
rs498636	11:64807541					3.11E-04	0.23	0.42	-542
rs6591869	11:64802267					9.49E-04	0.22	0.39	-5,816

a DNAm site identificator
b Position of the DNAm site respecting its annotated gene. TSS1500 refers a position 1500 bp to the gene promoter.
c mQTL correlation.
d SNP-DNAm site genetic distance.

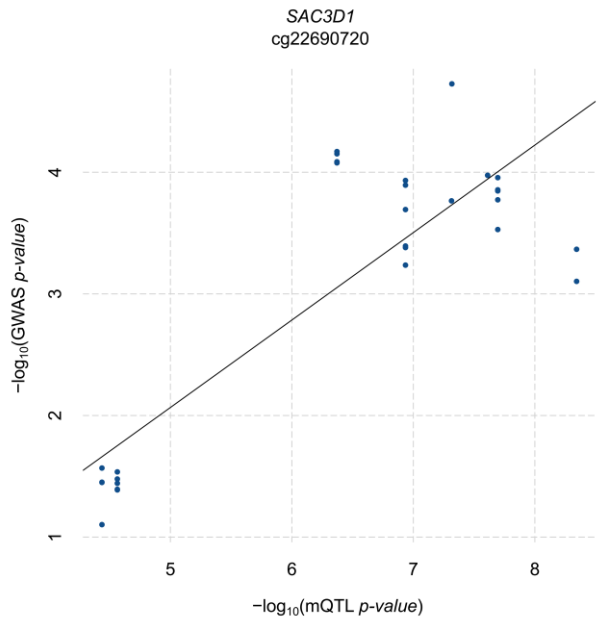

Figure 4.2. Comparison of the transformed significance levels (-$\log_{10}$ *p-values*) of the associated SNPs in the previous GWAS with those corresponding to their *cis*-mQTLs. The transformed *p-values* of the mQTL

and GWAS analyses were compared for those SNPs that form mQTL with the DNAm position cg22690720 located in the gene region of the *SAC3D1*.

Second, taking into account the total *cis*-mQTLs identified in the evaluated region, we found 268 interactions with significant methylation changes when comparing asymptomatic individuals and chronic Chagas cardiomyopathy patients. 89 (33%) DNAm sites were located in promotors (distance to TSS < 1500 bp), 175 (65%) in gene-bodies and 4 (1%) in intergenic regions (Table 4.3). The majority of these interactions affected a DNAm site located in the gene body of the DNA polymerase alpha 2, accessory subunit (*POLA2*) (Figure 4.3A) and also contain the most significant *cis*-mQTL in this locus (11:65063553−cg22229551, FDR=1.04x10^{-11}, R^2=0.72). Additionally nine SNPs from these *cis*-mQTLs were nominally associated with the disease in the previous GWAS (221).

Table 4.3. *cis* -mQTLs interactions that produce differentially methylation among chronic Chagas cardiomyopathy patients and asymptomatic individuals.

SNP ID	DNAm site ID[a]	Distance (bp)[b]	DNAm site function[c]	DNAm site gene	mQTL *p-value*	mQTL FDR	Beta	R^{2d}
rs72922019	cg00022866	-104203	Gene body	*CCDC88B*	0.0000158	0.0189	-0.37	0.3
rs1111934	cg14293362	450196	TSS200	*TM7SF2*	0.0000335	0.0338	-0.3	0.28
rs12292693	cg19940438	-146	TSS1500	*SPDYC*	1.24E-10	0.000000775	0.67	0.55
rs12292693	cg16849481	-260	TSS1500	*SPDYC*	2.93E-08	0.0000943	0.47	0.45
rs484147	cg22229551	-241058	Gene body	*POLA2*	0.00000878	0.0118	-0.3	0.32
rs61884708	cg26579892	478276	Gene body	*PLA2G16*	0.0000048	0.00722	-0.6	0.33
rs507062	cg23228688	69522	Intergenic	-	0.0000201	0.0228	-0.71	0.3

a DNAm site identificator
b SNP-DNAm site genetic distance.
c Position of the DNAm site respecting its annotated gene. TSS1500 and TSS200 refers a position at 1500-200 or less than 200 bp to the gene promoter.
d mQTL correlation.

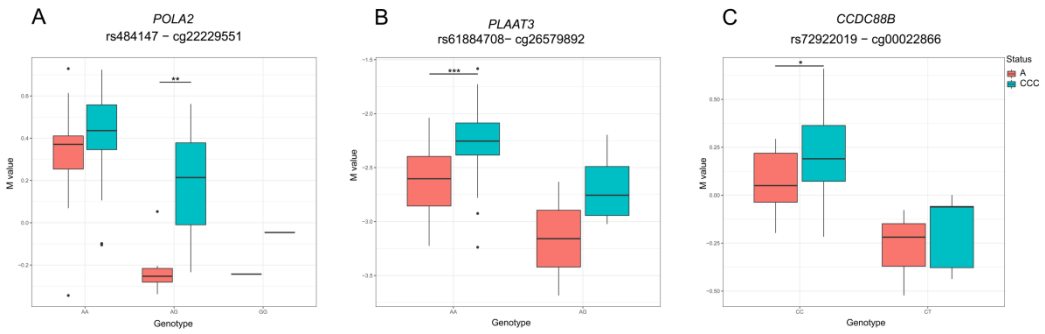

Figure 4.3. Box plots of three of the most interesting genes comparing DNAm levels between chronic Chagas cardiomyopathy patients (CCC) and asymptomatic individuals (A). Comparison of individuals' SNP genotypes and CpGs Mvalues for the most interesting mQTL genes that also showed differences in methylation patterns in cases and controls. Significant differences between genotypes are marked with symbols according to their level of significance calculated with an ANOVA test (* *p-value*≤0.05; ** *p-value*≤0.01; *** *p-value*≤0.001). Genotypes with no significant differences do not have a symbol.

The *cis*-mQTL with the most significant variation in the methylation patterns among cases and controls (*p-value*=8.09x10^{-04}), corresponded to the DNAm site within in the gene body of the Phospholipase A and acyltransferase 3 (*PLAAT3*) gene (rs61884708–cg26579892, FDR=7.22x10^{-03}, beta=0.12, R^2=0.33, Figure 4.3B). Among these results it is interesting also to highlight the interactions with the DNAm site located in the coiled-coil domain containing 88B (*CCDC88B*) gene body (rs72922019 – cg00022866, FDR=1.89x10^{-02}, beta=0.08, R^2=0.30, Figure 4.3C). This gene was previously reported to be differentially methylated and differentially expressed when comparing biopsies from heart tissue of chronic Chagas cardiomyopathy

patients and healthy donors (252), and we were able to corroborate these findings in an independent population.

4.2.2. Functional annotation analysis

To explore functional features of the genes related to the identified DNAm sites, several databases and online tools were queried. The majority of the methylated positions in the region were located within gene bodies (Figure 4.4), and those *cis*-mQTLs that interacted with DNAm sites located in intergenic regions were removed from further functional analyses.

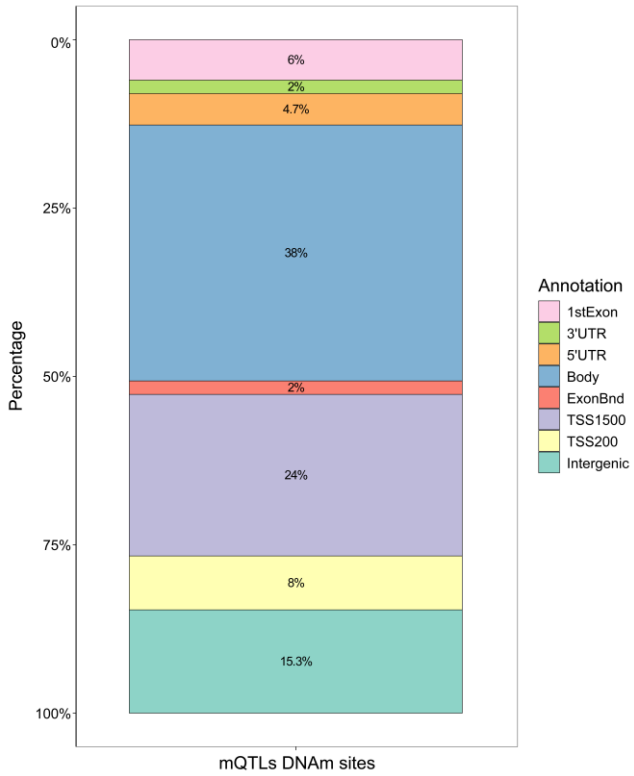

Figure 4.4. DNA methylation (DNAm) site positions respecting their nearest genes. For the 152 unique DNAm sites identified, their positions were compared with the Human Methylation EPIC manifest from Illumina and were expressed in percentages. According to their position, the DNAm sites can be located in the first exon (1st exon), in the 3' or 5' regulatory regions (3'UTR and 5'UTR), in the gene body (body), exons (exonBnd), at 1,500 or 200 to the transcription start site (TSS1500 and TSS200) or intergenic (intergenic).

According to the Open Targets Genetics database there are several association studies that have related the *POLA2* with hematological traits, such as mean corpuscular hemoglobin or red blood cell count (236, 253). One of these studies has also found an association of this locus with

cardiovascular disease (*p-value*=6x10^{-16}), highlighting the role of this gene in this trait (253). Additionally, different *loci* in the *CCDC88B* gene have been also associated with cardiovascular disease in the same dataset (*p-value*=1x10^{-10}) (253). Taking together, the *in silico* functional analysis of the resulting genes suggests the relationship of the associated region with different cardiovascular traits, supporting its relevance in the pathogenesis of the chronic Chagas cardiomyopathy.

4.3. Discussion

In this chapter we show the results of the first *cis*-mQTL analysis in Chagas disease patients from the Bolivian population, leveraging a previously reported genomic association, and its correlation with specific DNAm levels that may impact the regulation of gene expression. This assessment has provided further functional characterization of the associated locus, and has revealed significant interactions with relevant genes that have been previously related with several cardiovascular traits. As mentioned in previous chapters, the chronic cardiac form of the disease shows damage and inflammation of the myocardium as a response to the parasite, but also the microvascular system can be involved (254). Interestingly, these results revealed the regulation of gene expression of an inflammation related gene, validating previous findings, and highlighting the use of whole blood as a surrogate tissue for studying the chronic form of the disease. In addition, we identified regulatory mechanisms potentially involved on disease pathogenesis, exposing the functional implication of the previously associated *loci*.

Considering the sustained host-parasite interaction throughout human history, host genetic variation has been linked to the heterogeneous response observed to infectious agents (74). Nevertheless, functional studies are needed to determine the relevance of this genetic variation on disease pathogenesis (72). In this sense, epigenomic modifications are particularly ligated to pathological processes behind the chronic phases of infections, even when the first stages have been overcome (255, 256). Several of these modifications have been described in infectious diseases like viral (257), bacterial (258) and specifically, DNAm changes has been reported in parasitic diseases, such as Malaria, Leishmaniasis and Chagas disease (252, 259).

These results showed *cis*-mQTLs composed by SNPs in LD with the leading GWAS variant that were correlated with changes in the DNAm levels in the vicinity of the promoter of *SAC3D1*. This gene, as mentioned in chapter 2, is a transcriptional regulator of STAT5 (208), and increased phosphorylation levels of this protein have been associated with protection in human hearts (209). Among the significant interactions identified in the region, the top signal corresponded to a DNAm site located in the gene *PLCB3*, which encodes a member of the phospholipase C family. This protein family has been described to act in the induction of cardiac hypertrophy (260, 261). Moreover, another member of the phospholipase C family, the *PLCB2*, showed a significant correlation between methylation and gene expression levels in cardiac tissue, compared to healthy donors in a previous study in chronic Chagas cardiomyopathy (252). Although our results do not show significant differences on the methylation levels among asymptomatic and cardiomyopathic patients, we might speculate their potential role on the disease, based on their localization within the GWAS locus and their functional relevance.

From the total number of significant *cis*-mQTLs it is relevant to highlight those that produce substantial changes in methylation levels when comparing patients with chronic Chagas cardiomyopathy and asymptomatic individuals, since these interactions could have a greater implication on the regulation of gene expression in the pathogenesis of the disease. From these, the *CCDC88B* gene is of special interest because of its involvement in inflammatory conditions (262). This gene showed higher methylation levels within the gene body in cardiomyopathy patients, which is related with tissue-dependent regulation of gene expression (263). The *CCDC88B* has been associated with the most severe form of Malaria, cerebral malaria, where patients suffer inflammation of the brain tissue (264). In this study, the authors tested the role of the *CCDC88B* during this affection in a murine

model and showed that is a potential regulator of T cells function. This gene plays an important role in the inflammatory and host responses during parasitic infections (264, 265). This is of substantial interest since inflammation of the cardiac tissue also occurs during chronic Chagas cardiomyopathy (254, 266). In addition, *CCDC88B* has been also described in a murine model as relevant in the immune function of dendritic cells (262). In this sense, one subtype of dendritic cells, the tolerogenic dendritic cells, have been recently proposed to have therapeutic properties in a mouse model of chronic Chagas cardiomyopathy, given their role in the reduction of inflammation and fibrosis that contributes to reduce disease progression (44, 267). Interestingly, the *CCDC88B* was previously reported to suffer changes on its methylation and expression levels when cardiac tissues from patients with chronic Chagas cardiomyopathy and healthy donors were compared, showing also a correlation among both methylation an expression variations (252). The identification of this gene in whole blood from independent study samples highlights its role in the pathogenesis of the chronic Chagas cardiomyopathy, as well as emphasizes the use of whole blood as a surrogate tissue that allows the detection of disease-related molecules in a less invasive way. Thus, despite the high tissue specificity of the DNAm patterns (268), surrogate tissues as whole blood could be widely used, as in other diseases with high correlation of results (269).

The *PLAAT3* gene, also known as *PLA2G16*, showed the most significant variation in methylation levels, being also the chronic Chagas cardiomyopathy patients' methylation levels higher in this locus in comparison with asymptomatic individuals. This gene encodes an adipose-specific phospholipase whose expression has been reported to be reduced in patients with peripheral artery disease, which is a common affection in patients with coronary artery disease (270, 271). Thus, as mentioned for *PLCB3*, the role of the phospholipases during *T. cruzi* infection has been

widely studied as part of the lipid metabolism (272). Nevertheless, although this phospholipase seems to be related with other cardiovascular traits and exhibit higher expression in heart tissue, its potential relation with Chagas disease remains to be determined.

Finally, the most significant *cis*-mQTLs that produced variation in the methylation levels among cases and controls corresponded to DNAm sites located in the gene body of the *POLA2*. As for the other genes, *POLA2* showed higher methylation levels in chronic Chagas cardiomyopathy patients. Our *in silico* functional analyses related this gene with different hematological measurements in other association studies, in addition to be related with cardiovascular disease (236, 253). One of the hematological traits associated with *POLA2* is the mean corpuscular volume, which has been previously related with heart failure (273). In this sense, the relation of this gene with the pathogenesis of the chronic Chagas cardiomyopathy remains unclear, as gene expression information would be necessary to verify the direction of the reported methylation patterns. However, the identification of genes previously related to several cardiovascular traits and cardiac damage reinforces the role of the region under study previously associated with the disease (221). Therefore, our results suggest the DNAm as a possible driver of the variation observed in the differential development of the chronic Chagas cardiomyopathy.

This study has some limitations that are worth mentioning here. First, as the variation explained by *cis*-mQTLs is higher in blood (274), we focused our work in these interactions. Nevertheless, the addition of genetic distal interactions (i.e. *trans*-mQTLs) to the analysis could be also informative about other genes and pathways relevant for the disease. Interestingly, we used whole blood as a surrogate tissue and were able to validate the involvement of a previously reported gene in the cardiomyopathy. Thus,

further studies focused on the cardiac tissue would be interesting in order to validate the rest of the results. In order to confirm the methylation patterns associated with the identified interactions, validation assays such as pyrosequencing techniques would be of high interest in further studies focused in methylation patters. Another limitation is the analysis of variants with a MAF higher than 5%, which limits the number of low-frequency and rare variants evaluated in the region that might contribute to the disease risk as occurs in other complex diseases (246). Increasing the sample size may further allow the correlation of low frequency alleles with DNAm levels, and increase the statistical power in the differential methylation assessment. Finally, replication analyses would help to elucidate the generalization of these results to other Latin American populations, where the genetic variability of the parasite may contribute to different epigenetic modifications, as it has been previously reported for other parasitic infections (275).

This chapter describes the results of the first *cis*-mQTLs analysis carried out in chronic Chagas cardiomyopathy in patients from the Bolivian population. Numerous significant interactions between genetic variants and DNAm levels were identified within the only genomic region significantly associated with the disease. Several of these interactions showed also differential methylation in chronic Chagas cardiomyopathy patients, and validated previous findings. These results provide novel functional insights in the genetic background and gene regulation underlying the pathogenesis of the most severe form of Chagas disease.

GENERAL DISCUSSION

In the present thesis, different aspects of the host genetics in *T. cruzi* infection have been addressed given the nature of this complex infectious disease. Several type of genetic studies, candidate and GWAS, were described here highlighting the role of the genetic variation at different levels in Chagas disease. It is known that in addition to genomics, information provided form different types of data, like epigenomics as well as the molecular interaction among them, give a better understanding of the molecular complexity behind a disease (276).

The genetic associations described in our candidate genes studies pointed out, and confirmed, the involvement of the host genetic background in the immune response to the infection by the parasite. In addition, with the purpose of analyzing a great number of the genetic variability of the genome, a genome-wide analysis and meta-analysis of several Latin American populations were carried out, being the largest GWAS in Chagas disease to date. This analysis identified a genomic associated *locus* on chromosome 11, whose functional assessment was related with numerous genes previously associated with Chagas disease and other cardiac symptomatology such as *SAC3D1*, *BATF2* or *SNX15*. The absence of association at genomic level in one of the regions most related to the immune response, such as the HLA genes, may lead to the exclusion of this region as part of the genes related to the genetic contribution to the disease. Nevertheless, the complementary analysis of the local ancestry proportions in the Colombian population and their relationship to infection revealed the association of European and Native American ancestries with increased risk and protection to Chagas disease respectively. This association was observed in a region of chromosome 6 where the HLA and TRIM family genes are located. In addition, the selective sweep assessment highlighted several adaptive signals in this region functionally related with different HLA genes. In this way, we were able to observe an association in a region related to several

HLA genes, which highlights the relevance of these *loci* in the susceptibility to the infection, which could be driven by the selective pressure on this region. This leads us to believe that further studies focused on the analysis of SNPs located in the HLA region, as well as classical HLA alleles and amino acid positions, will allow us to have better knowledge of the role of this region in Chagas disease, as has been assessed in other infections such as Malaria, Tuberculosis or Leishmania (277-279). Occasionally, the absence of association in genomic regions such as this one may be due to the type of variant evaluated. Thus, structural variations such as CNVs may explain part of the variability in this region for Chagas disease (280). This structural variation has been assessed in other infectious diseases such as Malaria and HIV-1 infection (281). In the last one, it could be observed how the CNVs of the *KIR3DS1* gene, in combination with a specific HLA-B allele, showed a relation with the HIV-1 viral load and leaded to a slower progression to AIDS, being an example of the implication of the HLA in the disease identified by structural variations (281). Another important region where the effects of structural variations are key in the immune response is the complement system. This is an important component of the innate immune response that is also part of the first line of defense in infections by microorganisms (282). CNVs located in this region involve changes in the enzymatic cascade whose relationship with predisposition to autoimmune diseases is known (283). In the case of Chagas disease, it is known that during the acute phase there is an interaction between the parasite and different components of the complement system that leads to their activation in order to internalize the parasite. However, the enzymatic cascade is interrupted allowing the entry of trypanomastigotes into the cells and originating the chronic phase of the disease (282). Translating this interaction into the genetic context, one study evaluated the relationship among the disease with SNPs located in the gene region of one complement elements, the CR1 protein (284). In this study,

associations with the disease were found in several SNPs located in the gene coding for the CR1 protein, which the authors relate to an amino acid change that may affect the functionality of the protein and leads to the parasite persistence common in chronic Chagas disease patients (284). In this regard, although previous studies have evaluated the interaction of *T. cruzi* with the different complement elements, future analyses of structural variation in these characteristic regions would provide insight into the molecular mechanisms that lead to the activation or inactivation of this system during the acute and chronic phases of the disease.

Despite the efforts to apply genetic studies to underrepresented populations such as the Latin American ones, the limited number of references makes it difficult to assign functional outcomes to these populations, so they must be hypothesized. In addition, standard genomic references may leave out uncommon variants present in other populations. This was demonstrated in a sample of individuals of African descent, where it was observed that a considerable amount of sequences from the African genome were missing from the reference genome (285). These differences in genetic architecture between populations are also evident in the replicability of genetic studies in underrepresented populations.

In this sense, during 2018 the GWAS carried out in Latin American populations represent slightly more than 5% (286). This percentage will be maintained during 2021, being also comparable to the number of individuals included in this type of studies (Figure 6) (287). Thus, these events suggest that more population specific reference genomes are needed for population-based studies of human genetics.

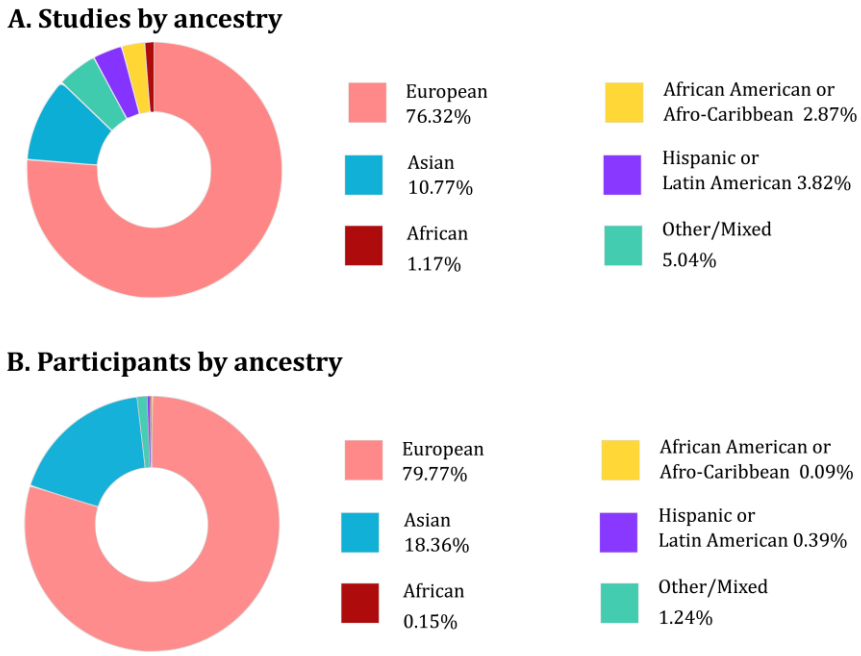

Figure 6. Number of genome-wide association studies and participants included in them in the discovery stages classified by ancestry during 2021 (Modified from https://gwasdiversitymonitor.com/).

The underrepresentation of Latin American populations in the reference panels is also reflected in functional analyses. In the case of epigenomics, as in their integration with other omics sciences, it is frequent also to use public databases in which, in most cases, the reference data used belong to the European population. This was especially relevant when it came to functionally characterizing the mQTLs identified in the associated region of chromosome 11. Regarding this, an important aspect to take into account in methylation studies is the obtaining of relevant tissue for the disease under study due to, like other epigenetic modifications, it is considered tissue-specific (288). This is remarkable in the case of tissues with limited access

such as the cardiac one and, as mentioned in previous chapters, the use of surrogate markers from tissues that are more accessible allow us to assess the epigenetic background behind a disease. Here, whole blood has been demonstrated to be a good candidate reporter of the influence of methylation changes in genomic variation in the context of the chronic Chagas cardiomyopathy. This aspect is also relevant not only considering the access to these tissues but also in terms of sample collection. The access to certain endemic regions and the cultural component in occasions can influence in the representation of ethnic minority groups (289). In this sense, it is important to take into consideration that some endemic areas have a difficult access and the risk that supposes the sample collection of an infectious disease. This limits the obtaining of larger sample sizes required for genome-wide association studies, which affects their statistical power. Therefore, it is noteworthy the observation of an association at the genomic level in the meta-analysis, despite the sample was limited and heterogeneous. Here, another important aspect that was necessary to take into consideration was the genetic heterogeneity present among and within Latin American populations. This variety is driven by the different admixture proportions of the European, Native American and African ancestries along and within the different Latin American countries (290).

The use of different host-based genetic approaches allowed us to better understand the pathogenesis of this disease and, ultimately, can lead to improve treatment, outcomes and the development of new drugs and vaccines. In this regard, a first step lies in the assessment of existing drugs that can be reused to combat the disease, which is known as drug repurposing. This approach allows optimizing the treatments development reducing time and cost of their production for which it is necessary to take into account the patient's genetic background in the context of the disease to be treated (291). In the case of parasitic diseases, these drugs should be

focused on controlling parasitemia, which is difficult to generalize among different parasites. Despite this, recent studies have evaluated the trypanocidal activity of some drugs such as nimesulide, an anti-inflammatory previously used in cancer and diabetes (292). In this sense, another interesting approach in the case of this disease would be focused on the chronic form and the inflammation that occurs in it, for example through the use of drugs used in other myocarditis. Therefore, the identification of common genetic biomarkers is an excellent tool in the knowledge of individual susceptibility to the disease and to be able to apply preventive measures.

FUTURE DIRECTIONS

The results widely discussed here are a novel contribution to the host genetic field in Chagas disease, although as mentioned, the assessment of several types of biomarkers is crucial for the knowledge of the molecular pathways involved. In this sense, our group recently obtained methylation data of individuals with chronic Chagas cardiomyopathy classified according to the severity of the disease using the Kuschnir scale. These methylation patterns were analyzed in 57 individuals belonging to the Bolivian cohort that also were used in the mQTL analysis described in chapter 4. Preliminary analyses of these data showed differences in the methylation patterns at a differentially methylated positions (DMPs) level when comparing the groups according to their severity (Figure 7). In this way, two main clusters can be observed showing a tendency to hypomethylation and hypermethylation, respectively, depending on the severity of the disease. These results are indicative of the existence of genetic variation in chronic Chagas cardiomyopathy patients beyond the variation in the DNA sequence.

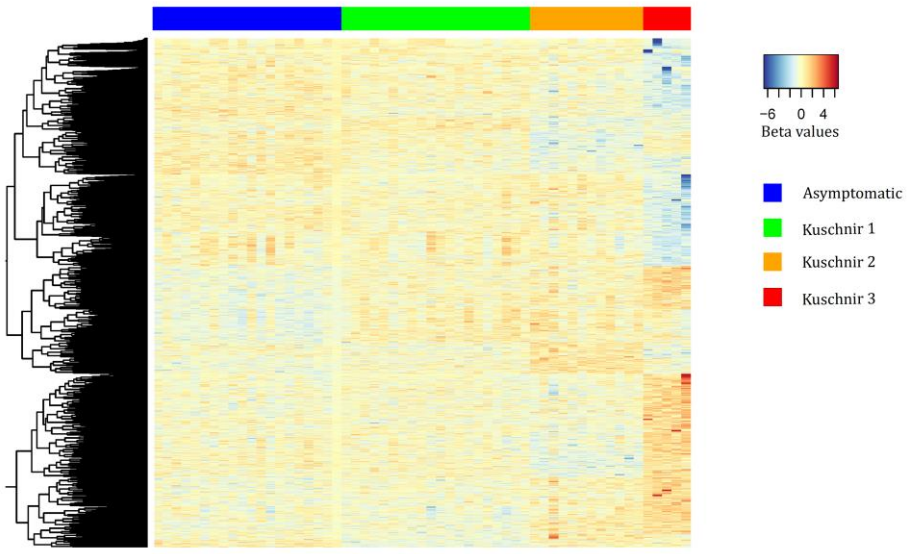

Figure 7. Heatmap of the DMPs (*p-value*<0.05) of 57 seropositive individuals from the bolivian cohort classified according to the Kuschnir scale. Hypermethylated CpGs

have higher beta values (red), while hypomethylated positions display lower beta values (blue). The rows repressent each CpG positon and the columns reffers the individuals.

Additionally, the analysis of differentially methylated regions (DMRs) in these patients is also ongoing. Preliminary results when comparing asymptomatic individuals with severe chronic Chagas cardiomyopathy patients (Kuschnir 2 and 3) pointed out two significant DMRs nearby the genes *DUSP22* and *VTRNA2-1*, which encode a phosphatase and a RNA polymerase III transcript. In this sense, protein phosphorylation is important for cellular signaling and, interestingly, the implication of the deregulation of these proteins in diseases such as cardiomyopathies or inflammatory conditions is well-known (293). Further *in silico* functional analyses of significant DMPs and DMRs will be needed in order to characterize these regions and their relation with the disease. Therefore, the comparison of these results with previous similar analyses carried out in the disease would clarify common molecular pathways when using different tissues.

Taking into account the relationship between DNA methylation and gene expression, the analysis of RNA-sequencing (RNA-seq) data will allows us to identify which genes are being expressed in a pathological process. In the case of Chagas disease, the latest study of the gene expression patterns was carried out in combination with a methylation analysis in a Brazilian population (252). In this sense, differences between Latin American populations can also be extrapolated to other genetic studies and, as a consequence, the expression patterns may also vary between populations (294, 295). Thus, the analysis of gene expression patterns in the Bolivian population would allow us to assess the differences and similarities of the pathways involved in the chronic cardiac form of the disease among different

populations. To this end, the gene expression profile of part of the Bolivian cohort included in the previous genetic analyses has been determined using the RNA-seq technology, and the association analysis with the disease and its integration with genetic and methylation data are being carried out. After this, further steps of this work will be focused in the functional characterization of these results and their enrichment in specific molecular pathways that may be related to the pathogenesis of the disease, which is also currently being carried out.

In general, these methodologies provide a wide amount of information about the biology of the disease that will help in the development of predictive models of disease risk, such as genomic risk score (GRS) studies, for the identification of markers of disease prognosis as well as therapeutic targets, especially focused on chronic Chagas cardiomyopathy. With the efforts behind this doctoral thesis we hope to have contributed to the knowledge of this neglected tropical disease.

CONCLUSIONS

1. The candidate gene studies performed confirmed the role of *IL17A* and *IL18*, but not *IL6* polymorphisms in Chagas disease. These results expose the activation of the host immune response at genetic level during initial stages of Chagas disease.

2. The largest GWAS in Chagas disease, and the meta-analysis of different Latin American populations, revealed a novel genome-wide statistically significant association with the development of chronic Chagas cardiomyopathy. This association was located in an intronic region of the *NAALADL1* gene and the vicinity of the *SAC3D1* and *SNX15* genes.

3. The functional characterization of the GWAS-identified novel locus using reference databases, revealed its relation with hypertension and blood pressure traits in PheWAS data. In addition, the variant was identified as eQTL in cardiac tissue for the *SNX15* gene, and further chromatin structure database showed its interaction with the promoter of *BATF2.*

4. The admixture mapping analysis revealed a protective association of the local Native American ancestry with the susceptibility to *T. cruzi* infection. This signal is located in the major histocompatibility complex region, where the TRIM and HLA gene families stood out for their functionality in the immune response against pathogens.

5. The mQTL analysis within the GWAS associated locus, allowed the identification of genes related to hematological and cardiac traits, as well as previously related to chronic Chagas cardiomyopathy, such as *POLA2*, *PLAAT3* and *CCDC88B*.

CONCLUSIONES

1. Los estudios de genes candidatos confirmaron el papel de las variantes de los genes *IL17A* e *IL18* en la enfermedad de Chagas. Sin embargo, la asociación con la enfermedad del gen *IL6* fue descartada. Estos resultados resaltaron la activación de la respuesta inmune del hospedador a nivel genético durante la fase inicial de la enfermedad.

2. El GWAS y meta-análisis de mayor tamaño muestral llevado a cabo hasta la fecha en la enfermedad de Chagas, reveló una asociación estadísticamente significativa a nivel genómico con el desarrollo de la cardiomiopatía Chagásica crónica. Esta asociación está localizada en un intrón del gen *NAALADL1* y en las inmediaciones de los genes *SAC3D1* y *SNX15*.

3. La caracterización funcional de la variante asociada en el GWAS reveló su relación con rasgos cardiovasculares. Además, fue identificada como eQTL en tejido cardíaco para el gen *SNX15* en bases de datos de referencia y que interacciona con el promotor de *BATF2* en datos de estructura de la cromatina.

4. El estudio de mapeo por mezcla reveló una asociación de protección de la ascendencia local Nativo Americana con la susceptibilidad a la infección por *T. cruzi.* Esta señal se localiza en la región del complejo mayor de histocompatibilidad, donde destaca la funcionalidad de las familias de genes TRIM y HLA en la respuesta inmune frente a patógenos.

5. El estudio de mQTLs en el locus asociado en el GWAS permitió la identificación de genes relacionados con caracteres hematológicos, cardíacos y previamente relacionados con la cardiomiopatía chagásica crónica, como el *POLA2*, *PLAAT3* y *CCDC88B*.

BIBLIOGRAPHY

1. Nunes MC, Dones W, Morillo CA, Encina JJ, Ribeiro AL, Council on Chagas Disease of the Interamerican Society of C. Chagas disease: an overview of clinical and epidemiological aspects. Journal of the American College of Cardiology. 2013;62(9):767-76.

2. Colombo V, Giacomelli A, Casazza G, Galimberti L, Bonazzetti C, Sabaini F, et al. Trypanosoma cruzi infection in Latin American pregnant women living outside endemic countries and frequency of congenital transmission: a systematic review and meta-analysis. Journal of travel medicine. 2021;28(1).

3. Lidani KCF, Andrade FA, Bavia L, Damasceno FS, Beltrame MH, Messias-Reason IJ, et al. Chagas Disease: From Discovery to a Worldwide Health Problem. Front Public Health. 2019;7:166.

4. Rassi A, Jr., Rassi A, Marin-Neto JA. Chagas disease. Lancet. 2010;375(9723):1388-402.

5. Aufderheide AC, Salo W, Madden M, Streitz J, Buikstra J, Guhl F, et al. A 9,000-year record of Chagas' disease. Proc Natl Acad Sci U S A. 2004;101(7):2034-9.

6. Coura JR. The main sceneries of Chagas disease transmission. The vectors, blood and oral transmissions--a comprehensive review. Mem Inst Oswaldo Cruz. 2015;110(3):277-82.

7. Angheben A, Boix L, Buonfrate D, Gobbi F, Bisoffi Z, Pupella S, et al. Chagas disease and transfusion medicine: a perspective from non-endemic countries. Blood transfusion = Trasfusione del sangue. 2015;13(4):540-50.

8. Silva-Dos-Santos D, Barreto-de-Albuquerque J, Guerra B, Moreira OC, Berbert LR, Ramos MT, et al. Unraveling Chagas disease transmission through the oral route: Gateways to Trypanosoma cruzi infection and target tissues. PLoS Negl Trop Dis. 2017;11(4):e0005507.

9. Perez-Molina JA, Molina I. Chagas disease. Lancet. 2018;391(10115):82-94.

10. Lee BY, Bacon KM, Bottazzi ME, Hotez PJ. Global economic burden of Chagas disease: a computational simulation model. The Lancet Infectious diseases. 2013;13(4):342-8.

11. Nagajyothi F, Machado FS, Burleigh BA, Jelicks LA, Scherer PE, Mukherjee S, et al. Mechanisms of Trypanosoma cruzi persistence in Chagas disease. Cellular microbiology. 2012;14(5):634-43.

12. Moretti NS, Mortara RA, Schenkman S. Trypanosoma cruzi. Trends in parasitology. 2020;36(4):404-5.

13. Zingales B. Trypanosoma cruzi genetic diversity: Something new for something known about Chagas disease manifestations, serodiagnosis and drug sensitivity. Acta Trop. 2018;184:38-52.

14. Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, Teixeira MM, et al. The revised Trypanosoma cruzi subspecific nomenclature: rationale, epidemiological relevance and research applications. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2012;12(2):240-53.

15. Jimenez V, Henriquez M, Galanti N, Riquelme G. Electrophysiological characterization of potassium conductive pathways in Trypanosoma cruzi. Journal of cellular biochemistry. 2011;112(4):1093-102.

16. Benaim G, Paniz-Mondolfi AE, Sordillo EM, Martinez-Sotillo N. Disruption of Intracellular Calcium Homeostasis as a Therapeutic Target Against Trypanosoma cruzi. Frontiers in cellular and infection microbiology. 2020;10:46.

17. Barrera P, Skorka C, Boktor M, Dave N, Jimenez V. A Novel Calcium-Activated Potassium Channel Controls Membrane Potential and Intracellular pH in Trypanosoma cruzi. Frontiers in cellular and infection microbiology. 2019;9:464.

18. Rodriguez-Duran J, Pinto-Martinez A, Castillo C, Benaim G. Identification and electrophysiological properties of a sphingosine-dependent plasma membrane Ca(2+) channel in Trypanosoma cruzi. The FEBS journal. 2019;286(19):3909-25.

19. Gutierrez FR, Mineo TW, Pavanelli WR, Guedes PM, Silva JS. The effects of nitric oxide on the immune system during Trypanosoma cruzi infection. Mem Inst Oswaldo Cruz. 2009;104 Suppl 1:236-45.

20. Teixeira AR, Hecht MM, Guimaro MC, Sousa AO, Nitz N. Pathogenesis of chagas' disease: parasite persistence and autoimmunity. Clinical microbiology reviews. 2011;24(3):592-630.

21. Pino-Marin A, Medina-Rincon GJ, Gallo-Bernal S, Duran-Crane A, Arango Duque AI, Rodriguez MJ, et al. Chagas Cardiomyopathy: From Romana Sign to Heart Failure and Sudden Cardiac Death. Pathogens. 2021;10(5).

22. Viotti R, Vigliano C, Lococo B, Bertocchi G, Petti M, Alvarez MG, et al. Long-term cardiac outcomes of treating chronic Chagas disease with benznidazole versus no treatment: a nonrandomized trial. Annals of internal medicine. 2006;144(10):724-34.

23. Norman FF, Lopez-Velez R. Chagas disease: comments on the 2018 PAHO Guidelines for diagnosis and management. Journal of travel medicine. 2019;26(7).

24. Bern C. Chagas' Disease. N Engl J Med. 2015;373(5):456-66.

25. Echavarria NG, Echeverria LE, Stewart M, Gallego C, Saldarriaga C. Chagas Disease: Chronic Chagas Cardiomyopathy. Current problems in cardiology. 2021;46(3):100507.

26. Rassi A, Jr., Rassi SG, Rassi A. Sudden death in Chagas' disease. Arq Bras Cardiol. 2001;76(1):75-96.

27. Rassi A, Jr., Rassi A, Little WC. Chagas' heart disease. Clinical cardiology. 2000;23(12):883-9.

28. Andrade JP, Marin Neto JA, Paola AA, Vilas-Boas F, Oliveira GM, Bacal F, et al. I Latin American Guidelines for the diagnosis and treatment of Chagas' heart disease: executive summary. Arq Bras Cardiol. 2011;96(6):434-42.

29. Kuschnir E, Sgammini H, Castro R, Evequoz C, Ledesma R, Brunetto J. [Evaluation of cardiac function by radioisotopic angiography, in patients with chronic Chagas cardiopathy]. Arq Bras Cardiol. 1985;45(4):249-56.

30. Carrasco HA, Barboza JS, Inglessis G, Fuenmayor A, Molina C. Left ventricular cineangiography in Chagas' disease: detection of early myocardial damage. American heart journal. 1982;104(3):595-602.

31. Acquatella H. Echocardiography in Chagas heart disease. Circulation. 2007;115(9):1124-31.

32. Pinazo MJ, Lacima G, Elizalde JI, Posada EJ, Gimeno F, Aldasoro E, et al. Characterization of digestive involvement in patients with chronic T. cruzi infection in Barcelona, Spain. PLoS Negl Trop Dis. 2014;8(8):e3105.

33. Matsuda NM, Miller SM, Evora PR. The chronic gastrointestinal manifestations of Chagas disease. Clinics. 2009;64(12):1219-24.

34. Bizai ML, Romina P, Antonela S, Olivera LV, Arias EE, Josefina DC, et al. Geographic distribution of Trypanosoma cruzi genotypes detected in chronic infected people from Argentina. Association with climatic variables and clinical manifestations of Chagas disease. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2020;78:104128.

35. Bahia MT, Diniz Lde F, Mosqueira VC. Therapeutical approaches under investigation for treatment of Chagas disease. Expert opinion on investigational drugs. 2014;23(9):1225-37.

36. Acosta-Herrera M, Strauss M, Casares-Marfil D, Martin J, Chagas Genetics CN. Genomic medicine in Chagas disease. Acta Trop. 2019;197:105062.

37. Morrot A, Villar SR, Gonzalez FB, Perez AR. Evasion and Immuno-Endocrine Regulation in Parasite Infection: Two Sides of the Same Coin in Chagas Disease? Frontiers in microbiology. 2016;7:704.

38. Machado FS, Dutra WO, Esper L, Gollob KJ, Teixeira MM, Factor SM, et al. Current understanding of immunity to Trypanosoma cruzi infection and pathogenesis of Chagas disease. Seminars in immunopathology. 2012;34(6):753-70.

39. Acevedo GR, Girard MC, Gomez KA. The Unsolved Jigsaw Puzzle of the Immune Response in Chagas Disease. Front Immunol. 2018;9:1929.

40. Caetano BC, Carmo BB, Melo MB, Cerny A, dos Santos SL, Bartholomeu DC, et al. Requirement of UNC93B1 reveals a critical role for TLR7 in host resistance to primary infection with Trypanosoma cruzi. Journal of immunology. 2011;187(4):1903-11.

41. Poveda C, Fresno M, Girones N, Martins-Filho OA, Ramirez JD, Santi-Rocca J, et al. Cytokine profiling in Chagas disease: towards understanding the association with infecting Trypanosoma cruzi discrete typing units (a BENEFIT TRIAL sub-study). PLoS One. 2014;9(3):e91154.

42. Cunha-Neto E, Nogueira LG, Teixeira PC, Ramasawmy R, Drigo SA, Goldberg AC, et al. Immunological and non-immunological effects of cytokines and chemokines in the pathogenesis of chronic Chagas disease cardiomyopathy. Mem Inst Oswaldo Cruz. 2009;104 Suppl 1:252-8.

43. Gil-Jaramillo N, Motta FN, Favali CB, Bastos IM, Santana JM. Dendritic Cells: A Double-Edged Sword in Immune Responses during Chagas Disease. Frontiers in microbiology. 2016;7:1076.

44. Santos ES, de Aragao-Franca LS, Meira CS, Cerqueira JV, Vasconcelos JF, Nonaka CKV, et al. Tolerogenic Dendritic Cells Reduce Cardiac Inflammation and Fibrosis in Chronic Chagas Disease. Front Immunol. 2020;11:488.

45. Kumar S, Tarleton RL. The relative contribution of antibody production and CD8+ T cell function to immune control of Trypanosoma cruzi. Parasite immunology. 1998;20(5):207-16.

46. Vitelli-Avelar DM, Sathler-Avelar R, Massara RL, Borges JD, Lage PS, Lana M, et al. Are increased frequency of macrophage-like and natural killer (NK) cells, together with high levels of NKT and CD4+CD25high T cells balancing activated CD8+ T cells, the key to control Chagas' disease morbidity? Clinical and experimental immunology. 2006;145(1):81-92.

47. Cardillo F, Postol E, Nihei J, Aroeira LS, Nomizo A, Mengel J. B cells modulate T cells so as to favour T helper type 1 and CD8+ T-cell responses in the acute phase of Trypanosoma cruzi infection. Immunology. 2007;122(4):584-95.

48. Rosser EC, Mauri C. Regulatory B cells: origin, phenotype, and function. Immunity. 2015;42(4):607-12.

49. Dutra WO, Gollob KJ. Current concepts in immunoregulation and pathology of human Chagas disease. Current opinion in infectious diseases. 2008;21(3):287-92.

50. De Bona E, Lidani KCF, Bavia L, Omidian Z, Gremski LH, Sandri TL, et al. Autoimmunity in Chronic Chagas Disease: A Road of Multiple Pathways to Cardiomyopathy? Front Immunol. 2018;9:1842.

51. Labovsky V, Smulski CR, Gomez K, Levy G, Levin MJ. Anti-beta1-adrenergic receptor autoantibodies in patients with chronic Chagas heart disease. Clinical and experimental immunology. 2007;148(3):440-9.

52. Tarleton RL. Chagas disease: a role for autoimmunity? Trends in parasitology. 2003;19(10):447-51.

53. Ku CS, Loy EY, Salim A, Pawitan Y, Chia KS. The discovery of human genetic variations and their use as disease markers: past, present and future. Journal of human genetics. 2010;55(7):403-15.

54. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-65.

55. Lewis CM. Genetic association studies: design, analysis and interpretation. Briefings in bioinformatics. 2002;3(2):146-53.

56. International Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931-45.

57. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.

58. Patnala R, Clements J, Batra J. Candidate gene association studies: a comprehensive guide to useful in silico tools. BMC genetics. 2013;14:39.

59. Jorgensen TJ, Ruczinski I, Kessing B, Smith MW, Shugart YY, Alberg AJ. Hypothesis-driven candidate gene association studies: practical design and analytical considerations. American journal of epidemiology. 2009;170(8):986-93.

60. International HapMap C. The International HapMap Project. Nature. 2003;426(6968):789-96.

61. Verlouw JAM, Clemens E, de Vries JH, Zolk O, Verkerk A, Am Zehnhoff-Dinnesen A, et al. A comparison of genotyping arrays. European journal of human genetics : EJHG. 2021;29(11):1611-24.

62. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. 2017;101(1):5-22.

63. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. Nature Reviews Methods Primers. 2021;1(1):59.

64. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genetic epidemiology. 2008;32(4):381-5.

65. Norris ET, Wang L, Conley AB, Rishishwar L, Marino-Ramirez L, Valderrama-Aguirre A, et al. Genetic ancestry, admixture and health determinants in Latin America. BMC genomics. 2018;19(Suppl 8):861.

66. Grinde KE, Brown LA, Reiner AP, Thornton TA, Browning SR. Genome-wide Significance Thresholds for Admixture Mapping Studies. Am J Hum Genet. 2019;104(3):454-65.

67. Rishishwar L, Conley AB, Wigington CH, Wang L, Valderrama-Aguirre A, Jordan IK. Ancestry, admixture and fitness in Colombian genomes. Sci Rep. 2015;5:12376.

68. Norris ET, Rishishwar L, Chande AT, Conley AB, Ye K, Valderrama-Aguirre A, et al. Admixture-enabled selection for rapid adaptive evolution in the Americas. Genome biology. 2020;21(1):29.

69. Shriner D. Overview of Admixture Mapping. Current protocols in human genetics. 2017;94:1 23 1-1 8.

70. Gignoux CR, Torgerson DG, Pino-Yanes M, Uricchio LH, Galanter J, Roth LA, et al. An admixture mapping meta-analysis implicates genetic variation at 18q21 with asthma susceptibility in Latinos. The Journal of allergy and clinical immunology. 2019;143(3):957-69.

71. Zhang L, Lu Q, Chang C. Epigenetics in Health and Disease. Advances in experimental medicine and biology. 2020;1253:3-55.

72. Villicana S, Bell JT. Genetic impacts on DNA methylation: research findings and future perspectives. Genome biology. 2021;22(1):127.

73. Jiang Z, Zhou X, Li R, Michal JJ, Zhang S, Dodson MV, et al. Whole transcriptome analysis with sequencing: methods, challenges and potential solutions. Cellular and molecular life sciences : CMLS. 2015;72(18):3425-39.

74. Kwok AJ, Mentzer A, Knight JC. Host genetics and infectious disease: new tools, insights and translational opportunities. Nat Rev Genet. 2021;22(3):137-53.

75. Quintana-Murci L. Human Immunology through the Lens of Evolutionary Genetics. Cell. 2019;177(1):184-99.

76. Jallow M, Teo YY, Small KS, Rockett KA, Deloukas P, Clark TG, et al. Genome-wide and fine-resolution association analysis of malaria in West Africa. Nat Genet. 2009;41(6):657-65.

77. Timmann C, Thye T, Vens M, Evans J, May J, Ehmen C, et al. Genome-wide association study indicates two novel resistance loci for severe malaria. Nature. 2012;489(7416):443-6.

78. Band G, Le QS, Jostins L, Pirinen M, Kivinen K, Jallow M, et al. Imputation-based meta-analysis of severe malaria in three African populations. PLoS genetics. 2013;9(5):e1003509.

79. Mangano VD, Modiano D. Host genetics and parasitic infections. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2014;20(12):1265-75.

80. Castellucci L, Jamieson SE, Miller EN, de Almeida LF, Oliveira J, Magalhaes A, et al. FLI1 polymorphism affects susceptibility to cutaneous leishmaniasis in Brazil. Genes Immun. 2011;12(7):589-94.

81. Castellucci L, Jamieson SE, Miller EN, Menezes E, Oliveira J, Magalhaes A, et al. CXCR1 and SLC11A1 polymorphisms affect susceptibility to cutaneous leishmaniasis in Brazil: a case-control and family-based study. BMC medical genetics. 2010;11:10.

82. Mohamed HS, Ibrahim ME, Miller EN, White JK, Cordell HJ, Howson JM, et al. SLC11A1 (formerly NRAMP1) and susceptibility to visceral leishmaniasis in The Sudan. European journal of human genetics : EJHG. 2004;12(1):66-74.

83. Castellucci LC, Almeida L, Cherlin S, Fakiola M, Francis RW, Carvalho EM, et al. A Genome-wide Association Study Identifies SERPINB10, CRLF3, STX7, LAMP3, IFNG-AS1, and KRT80 As Risk Loci Contributing to Cutaneous Leishmaniasis in Brazil. Clin Infect Dis. 2021;72(10):e515-e25.

84. Kabore JW, Ilboudo H, Noyes H, Camara O, Kabore J, Camara M, et al. Candidate gene polymorphisms study between human African trypanosomiasis clinical phenotypes in Guinea. PLoS Negl Trop Dis. 2017;11(8):e0005833.

85. Cruz-Tapias P, Agmon-Levin N, Israeli E, Anaya JM, Shoenfeld Y. Autoimmune (auto-inflammatory) syndrome induced by adjuvants (ASIA)--animal models as a proof of concept. Current medicinal chemistry. 2013;20(32):4030-6.

86. Ayo CM, Dalalio MM, Visentainer JE, Reis PG, Sippert EA, Jarduli LR, et al. Genetic susceptibility to Chagas disease: an overview about the infection and about the association between disease and the immune response genes. Biomed Res Int. 2013;2013:284729.

87. Ortega Zamora Y, Escamilla Rojas LJ, Villa Sandoval EM, Vela Porras JS, Cossio Contrera EY, Cubides Romero SS, et al. Chagas disease immunogenetics: elusive markers of disease progression. Expert review of cardiovascular therapy. 2017;15(5):367-76.

88. Nieto A, Beraun Y, Collado MD, Caballero A, Alonso A, Gonzalez A, et al. HLA haplotypes are associated with differential susceptibility to Trypanosoma cruzi infection. Tissue Antigens. 2000;55(3):195-8.

89. Cruz-Robles D, Reyes PA, Monteon-Padilla VM, Ortiz-Muniz AR, Vargas-Alarcon G. MHC class I and class II genes in Mexican patients with Chagas disease. Human immunology. 2004;65(1):60-5.

90. del Puerto F, Nishizawa JE, Kikuchi M, Roca Y, Avilas C, Gianella A, et al. Protective human leucocyte antigen haplotype, HLA-DRB1*01-B*14, against chronic Chagas disease in Bolivia. PLoS Negl Trop Dis. 2012;6(3):e1587.

91. Salvador F, Sanchez-Montalva A, Martinez-Gallo M, Sala-Cunill A, Vinas L, Garcia-Prat M, et al. Evaluation of cytokine profile and HLA association in benznidazole related cutaneous reactions in patients with Chagas disease. Clin Infect Dis. 2015;61(11):1688-94.

92. Ayo CM, Oliveira AP, Camargo AV, Mattos CC, Bestetti RB, Mattos LC. Association of the Functional MICA-129 Polymorphism With the Severity of Chronic Chagas Heart Disease. Clin Infect Dis. 2015;61(8):1310-3.

93. Dias FC, Mendes-Junior CT, Silva MC, Tristao FS, Dellalibera-Joviliano R, Moreau P, et al. Human leucocyte antigen-G (HLA-G) and its murine functional homolog Qa2 in the Trypanosoma cruzi Infection. Mediators Inflamm. 2015;2015:595829.

94. Sabbagh A, Sonon P, Sadissou I, Mendes-Junior CT, Garcia A, Donadi EA, et al. The role of HLA-G in parasitic diseases. Hla. 2018;91(4):255-70.

95. Dedoussis GV, Panagiotakos DB, Vidra NV, Louizou E, Chrysohoou C, Germanos A, et al. Association between TNF-alpha -308G>A polymorphism and the development of acute coronary syndromes in Greek subjects: the CARDIO2000-GENE Study. Genetics in medicine : official journal of the American College of Medical Genetics. 2005;7(6):411-6.

96. Zhang P, Wu X, Li G, He Q, Dai H, Ai C, et al. Tumor necrosis factor-alpha gene polymorphisms and susceptibility to ischemic heart disease: A systematic review and meta-analysis. Medicine. 2017;96(14):e6569.

97. Campelo V, Dantas RO, Simoes RT, Mendes-Junior CT, Sousa SM, Simoes AL, et al. TNF microsatellite alleles in Brazilian Chagasic patients. Digestive diseases and sciences. 2007;52(12):3334-9.

98. Criado L, Florez O, Martin J, Gonzalez CI. Genetic polymorphisms in TNFA/TNFR2 genes and Chagas disease in a Colombian endemic population. Cytokine. 2012;57(3):398-401.

99. Drigo SA, Cunha-Neto E, Ianni B, Cardoso MR, Braga PE, Fae KC, et al. TNF gene polymorphisms are associated with reduced survival in severe Chagas' disease cardiomyopathy patients. Microbes and infection. 2006;8(3):598-603.

100. Rodriguez-Perez JM, Cruz-Robles D, Hernandez-Pacheco G, Perez-Hernandez N, Murguia LE, Granados J, et al. Tumor necrosis factor-alpha promoter polymorphism in Mexican patients with Chagas' disease. Immunology letters. 2005;98(1):97-102.

101. Alvarado-Arnez LE, Batista AM, Alves SM, Melo G, Lorena VMB, Cardoso CC, et al. Single nucleotide polymorphisms of cytokine-related genes and association with clinical outcome in a Chagas disease case-control study from Brazil. Mem Inst Oswaldo Cruz. 2018;113(6):e170489.

102. Beraun Y, Nieto A, Collado MD, Gonzalez A, Martin J. Polymorphisms at tumor necrosis factor (TNF) loci are not associated with Chagas' disease. Tissue Antigens. 1998;52(1):81-3.

103. Drigo SA, Cunha-Neto E, Ianni B, Mady C, Fae KC, Buck P, et al. Lack of association of tumor necrosis factor-alpha polymorphisms with Chagas disease in Brazilian patients. Immunology letters. 2007;108(1):109-11.

104. Leon Rodriguez DA, Carmona FD, Gonzalez CI, Martin J. Evaluation of VDR gene polymorphisms in Trypanosoma cruzi infection and chronic Chagasic cardiomyopathy. Sci Rep. 2016;6:31263.

105. Leon Rodriguez DA, Gonzalez CI, Martin J. Analysis of association of FOXO3 gene with Trypanosoma cruzi infection and chronic Chagasic cardiomyopathy. Hla. 2016;87(6):449-52.

106. Leon Rodriguez DA, Echeverria LE, Gonzalez CI, Martin J. Investigation of the role of IL17A gene variants in Chagas disease. Genes Immun. 2015;16(8):536-40.

107. Torres OA, Calzada JE, Beraun Y, Morillo CA, Gonzalez A, Gonzalez CI, et al. Lack of association between IL-6-174G/C gene polymorphism and Chagas disease. Tissue Antigens. 2010;76(2):131-4.

108. Torres OA, Calzada JE, Beraun Y, Morillo CA, Gonzalez A, Gonzalez CI, et al. Role of the IFNG +874T/A polymorphism in Chagas disease in a Colombian population. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2010;10(5):682-5.

109. Nogueira LG, Frade AF, Ianni BM, Laugier L, Pissetti CW, Cabantous S, et al. Functional IL18 polymorphism and susceptibility to Chronic Chagas Disease. Cytokine. 2015;73(1):79-83.

110. de Oliveira AP, Ayo CM, Bestetti RB, Brandao de Mattos CC, Cavasini CE, de Mattos LC. The role of CCR5 in Chagas disease - a systematic review.

Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2016;45:132-7.

111. Juiz NA, Estupinan E, Hernandez D, Garcilazo A, Chadi R, Morales Sanfurgo G, et al. Association study between CCR2-CCR5 genes polymorphisms and chronic Chagas heart disease in Wichi and in admixed populations from Argentina. PLoS Negl Trop Dis. 2019;13(1):e0007033.

112. Machuca MA, Suarez EU, Echeverria LE, Martin J, Gonzalez CI. SNP/haplotype associations of CCR2 and CCR5 genes with severity of chagasic cardiomyopathy. Human immunology. 2014;75(12):1210-5.

113. Nogueira LG, Santos RH, Ianni BM, Fiorelli AI, Mairena EC, Benvenuti LA, et al. Myocardial chemokine expression and intensity of myocarditis in Chagas cardiomyopathy are controlled by polymorphisms in CXCL9 and CXCL10. PLoS Negl Trop Dis. 2012;6(10):e1867.

114. Torres OA, Calzada JE, Beraun Y, Morillo CA, Gonzalez CI, Gonzalez A, et al. Association of the macrophage migration inhibitory factor -173G/C polymorphism with Chagas disease. Human immunology. 2009;70(7):543-6.

115. Frade AF, Pissetti CW, Ianni BM, Saba B, Lin-Wang HT, Nogueira LG, et al. Genetic susceptibility to Chagas disease cardiomyopathy: involvement of several genes of the innate immunity and chemokine-dependent migration pathways. BMC infectious diseases. 2013;13:587.

116. Ramasawmy R, Cunha-Neto E, Fae KC, Borba SC, Teixeira PC, Ferreira SC, et al. Heterozygosity for the S180L variant of MAL/TIRAP, a gene expressing an adaptor protein in the Toll-like receptor pathway, is associated with lower risk of developing chronic Chagas cardiomyopathy. The Journal of infectious diseases. 2009;199(12):1838-45.

117. Jorge SE, Abreu CF, Guariento ME, Sonati Mde F. Haptoglobin genotypes in Chagas' disease. Clinical biochemistry. 2010;43(3):314-6.

118. Mundaray Fernandez N, Fernandez-Mestre M. The role of haptoglobin genotypes in Chagas disease. Disease markers. 2014;2014:793646.

119. Sadrzadeh SM, Bozorgmehr J. Haptoglobin phenotypes in health and disorders. American journal of clinical pathology. 2004;121 Suppl:S97-104.

120. Luz PR, Miyazaki MI, Neto NC, Nisihara RM, Messias-Reason IJ. High levels of mannose-binding lectin are associated with the risk of severe cardiomyopathy in chronic Chagas Disease. International journal of cardiology. 2010;143(3):448-50.

121. Weitzel T, Zulantay I, Danquah I, Hamann L, Schumann RR, Apt W, et al. Mannose-binding lectin and Toll-like receptor polymorphisms and Chagas disease in Chile. The American journal of tropical medicine and hygiene. 2012;86(2):229-32.

122. Leon Rodriguez DA, Carmona FD, Echeverria LE, Gonzalez CI, Martin J. IL18 Gene Variants Influence the Susceptibility to Chagas Disease. PLoS Negl Trop Dis. 2016;10(3):e0004583.

123. Reis PG, Ayo CM, de Mattos LC, Brandao de Mattos CC, Sakita KM, de Moraes AG, et al. Genetic Polymorphisms of IL17 and Chagas Disease in the South and Southeast of Brazil. Journal of immunology research. 2017;2017:1017621.

124. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21(2):263-5.

125. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

126. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am J Hum Genet. 2011;88(5):586-98.

127. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet. 2006;38(2):209-13.

128. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res. 2016;44(D1):D877-81.

129. Cheneby J, Gheorghe M, Artufel M, Mathelier A, Ballester B. ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. Nucleic Acids Res. 2018;46(D1):D267-D75.

130. Cunha-Neto E, Chevillard C. Chagas disease cardiomyopathy: immunopathology and genetics. Mediators Inflamm. 2014;2014:683230.

131. Zhang H, Wheeler W, Hyland PL, Yang Y, Shi J, Chatterjee N, et al. A Powerful Procedure for Pathway-Based Meta-analysis Using Summary Statistics Identifies 43 Pathways Associated with Type II Diabetes in European Populations. PLoS genetics. 2016;12(6):e1006122.

132. Lona-Durazo F, Hernandez-Pacheco N, Fan S, Zhang T, Choi J, Kovacs MA, et al. Meta-analysis of GWA studies provides new insights on the genetic architecture of skin pigmentation in recently admixed populations. BMC genetics. 2019;20(1):59.

133. Courtin D, Milet J, Jamonneau V, Yeminanga CS, Kumeso VK, Bilengue CM, et al. Association between human African trypanosomiasis and the IL6 gene in a Congolese population. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2007;7(1):60-8.

134. Wujcicka W, Gaj Z, Wilczynski J, Nowakowska D. Contribution of IL6 -174 G>C and IL1B +3954 C>T polymorphisms to congenital infection with Toxoplasma gondii. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology. 2015;34(11):2287-94.

135. Ahouty B, Koffi M, Ilboudo H, Simo G, Matovu E, Mulindwa J, et al. Candidate genes-based investigation of susceptibility to Human African Trypanosomiasis in Cote d'Ivoire. PLoS Negl Trop Dis. 2017;11(10):e0005992.

136. Almeida NCC, Queiroz MAF, Lima SS, Brasil Costa I, Ayin Fossa MA, Vallinoto ACR, et al. Association of Chlamydia trachomatis, C. pneumoniae, and IL-6 and IL-8 Gene Alterations With Heart Diseases. Front Immunol. 2019;10:87.

137. Chevillard C, Nunes JPS, Frade AF, Almeida RR, Pandey RP, Nascimento MS, et al. Disease Tolerance and Pathogen Resistance Genes May Underlie Trypanosoma cruzi Persistence and Differential Progression to Chagas Disease Cardiomyopathy. Front Immunol. 2018;9:2791.

138. Lopez L, Arai K, Gimenez E, Jimenez M, Pascuzo C, Rodriguez-Bonfante C, et al. [C-reactive protein and interleukin-6 serum levels increase as Chagas disease progresses towards cardiac failure]. Revista espanola de cardiologia. 2006;59(1):50-6.

139. Barbosa-Ferreira JM, Mady C, Ianni BM, Lopes HF, Ramires FJ, Salemi VM, et al. Dysregulation of Autonomic Nervous System in Chagas' Heart Disease Is Associated with Altered Adipocytokines Levels. PLoS One. 2015;10(7):e0131447.

140. Sanmarco LM, Visconti LM, Eberhardt N, Ramello MC, Ponce NE, Spitale NB, et al. IL-6 Improves the Nitric Oxide-Induced Cytotoxic CD8+ T Cell Dysfunction in Human Chagas Disease. Front Immunol. 2016;7:626.

141. Antunes D, Marins-Dos-Santos A, Ramos MT, Mascarenhas BAS, Moreira CJC, Farias-de-Oliveira DA, et al. Oral Route Driven Acute Trypanosoma cruzi Infection Unravels an IL-6 Dependent Hemostatic Derangement. Front Immunol. 2019;10:1073.

142. Aker S, Bantis C, Reis P, Kuhr N, Schwandt C, Grabensee B, et al. Influence of interleukin-6 G-174C gene polymorphism on coronary artery disease, cardiovascular complications and mortality in dialysis patients. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association. 2009;24(9):2847-51.

143. Ansari WM, Humphries SE, Naveed AK, Khan OJ, Khan DA. Influence of cytokine gene polymorphisms on proinflammatory/anti-inflammatory cytokine imbalance in premature coronary artery disease. Postgraduate medical journal. 2017;93(1098):209-14.

144. Chen L, Zhang Z, Huang J, Jin M. Association between rs1800795 polymorphism in the interleukin-6 gene and the risk of polycystic ovary syndrome: A meta-analysis. Medicine. 2018;97(29):e11558.

145. Schor D, Porto LC, Roma EH, Quintana MSB, Fabricio-Silva GM, Bonecini-Almeida MG, et al. Lack of association between single-nucleotide

polymorphisms of pro- and anti-inflammatory cytokines and HTLV-1-associated myelopathy / tropical spastic paraparesis development in patients from Rio de Janeiro, Brazil. BMC infectious diseases. 2018;18(1):593.

146. Fleury A, Alaez C, Dessein A, Rosetti M, Saenz B, Hernandez M, et al. No association of IL2, IL4, IL6, TNF, and IFNG gene polymorphisms was found with Taenia solium human infection or neurocysticercosis severity in a family-based study. Human immunology. 2018;79(7):578-82.

147. Georganas C, Liu H, Perlman H, Hoffmann A, Thimmapaya B, Pope RM. Regulation of IL-6 and IL-8 expression in rheumatoid arthritis synovial fibroblasts: the dominant role for NF-kappa B but not C/EBP beta or c-Jun. Journal of immunology. 2000;165(12):7199-206.

148. Iliopoulos D, Hirsch HA, Struhl K. An epigenetic switch involving NF-kappaB, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation. Cell. 2009;139(4):693-706.

149. Ishida K, Kobayashi T, Ito S, Komatsu Y, Yokoyama T, Okada M, et al. Interleukin-6 gene promoter methylation in rheumatoid arthritis and chronic periodontitis. Journal of periodontology. 2012;83(7):917-25.

150. Didion SP. Cellular and Oxidative Mechanisms Associated with Interleukin-6 Signaling in the Vasculature. Int J Mol Sci. 2017;18(12).

151. Morales-Nebreda L, McLafferty FS, Singer BD. DNA methylation as a transcriptional regulator of the immune system. Translational research : the journal of laboratory and clinical medicine. 2019;204:1-18.

152. Bryc K, Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Auton A, et al. Colloquium paper: genome-wide patterns of population structure and

admixture among Hispanic/Latino populations. Proc Natl Acad Sci U S A. 2010;107 Suppl 2:8954-61.

153. Bermejo DA, Jackson SW, Gorosito-Serran M, Acosta-Rodriguez EV, Amezcua-Vesely MC, Sather BD, et al. Trypanosoma cruzi trans-sialidase initiates a program independent of the transcription factors RORgammat and Ahr that leads to IL-17 production by activated B cells. Nature immunology. 2013;14(5):514-22.

154. Gorosito Serran M, Tosello Boari J, Fiocca Vernengo F, Beccaria CG, Ramello MC, Bermejo DA, et al. Unconventional Pro-inflammatory CD4(+) T Cell Response in B Cell-Deficient Mice Infected with Trypanosoma cruzi. Front Immunol. 2017;8:1548.

155. Miyazaki Y, Hamano S, Wang S, Shimanoe Y, Iwakura Y, Yoshida H. IL-17 is necessary for host protection against acute-phase Trypanosoma cruzi infection. Journal of immunology. 2010;185(2):1150-7.

156. Chen K, Kolls JK. Interluekin-17A (IL17A). Gene. 2017;614:8-14.

157. Taylor PR, Roy S, Leal SM, Jr., Sun Y, Howell SJ, Cobb BA, et al. Activation of neutrophils by autocrine IL-17A-IL-17RC interactions during fungal infection is regulated by IL-6, IL-23, RORgammat and dectin-2. Nature immunology. 2014;15(2):143-51.

158. Nordang GB, Viken MK, Hollis-Moffatt JE, Merriman TR, Forre OT, Helgetveit K, et al. Association analysis of the interleukin 17A gene in Caucasian rheumatoid arthritis patients from Norway and New Zealand. Rheumatology. 2009;48(4):367-70.

159. Duan Y, Shi JN, Pan C, Chen HL, Zhang SZ. Association between the interleukin-17A -197G>A (rs2275913) polymorphism and risk of digestive

cancer. Asian Pacific journal of cancer prevention : APJCP. 2014;15(21):9295-300.

160. Hayashi R, Tahara T, Shiroeda H, Saito T, Nakamura M, Tsutsumi M, et al. Influence of IL17A polymorphisms (rs2275913 and rs3748067) on the susceptibility to ulcerative colitis. Clinical and experimental medicine. 2013;13(4):239-44.

161. Pasha HF, Tantawy EA, Youssef MA. Osteopontin and interleukin-17A genes polymorphisms in Egyptian systemic lupus erythematosus patients: A relation to disease activity and severity. Gene. 2019;702:107-13.

162. Espinoza JL, Takami A, Nakata K, Onizuka M, Kawase T, Akiyama H, et al. A genetic variant in the IL-17 promoter is functionally associated with acute graft-versus-host disease after unrelated bone marrow transplantation. PLoS One. 2011;6(10):e26229.

163. Rolandelli A, Hernandez Del Pino RE, Pellegrini JM, Tateosian NL, Amiano NO, de la Barrera S, et al. The IL-17A rs2275913 single nucleotide polymorphism is associated with protection to tuberculosis but related to higher disease severity in Argentina. Sci Rep. 2017;7:40666.

164. Huang HT, Lu YL, Wang R, Qin HM, Wang CF, Wang JL, et al. The association of IL-17A polymorphisms with IL-17A serum levels and risk of ischemic stroke. Oncotarget. 2017;8(61):103499-508.

165. Li N, Zhu Q, Li Z, Han Q, Zhang G, Chen J, et al. IL17A gene polymorphisms, serum IL-17A and IgE levels, and hepatocellular carcinoma risk in patients with chronic hepatitis B virus infection. Molecular carcinogenesis. 2014;53(6):447-57.

166. Mohammadipour K, Mansouri R, Salmanpour R, Haghshenas MR, Erfani N. Investigation of Interleukin-17 Gene Polymorphisms and Serum Levels in

Patients with Basal Cell Carcinoma of the Skin. Iranian journal of immunology : IJI. 2019;16(1):53-61.

167. da Matta Guedes PM, Gutierrez FR, Maia FL, Milanezi CM, Silva GK, Pavanelli WR, et al. IL-17 produced during Trypanosoma cruzi infection plays a central role in regulating parasite-induced myocarditis. PLoS Negl Trop Dis. 2010;4(2):e604.

168. Almeida MS, Lorena VMB, Medeiros CA, Junior WO, Cavalcanti M, Martins SM, et al. Alternative Th17 and CD4(+) CD25(+) FoxP3(+) cell frequencies increase and correlate with worse cardiac function in Chagas cardiomyopathy. Scandinavian journal of immunology. 2018;87(4):e12650.

169. Bestetti RB, Dellalibera-Joviliano R, Lopes GS, Faria-Jr M, Furlan-Daniel R, Lopes KC, et al. Determination of the Th1, Th2, Th17, and Treg cytokine profile in patients with chronic Chagas heart disease and systemic arterial hypertension. Heart and vessels. 2019;34(1):123-33.

170. Amezcua Vesely MC, Rodriguez C, Gruppi A, Acosta Rodriguez EV. Interleukin-17 mediated immunity during infections with Trypanosoma cruzi and other protozoans. Biochimica et biophysica acta Molecular basis of disease. 2020;1866(5):165706.

171. Nogueira LG, Santos RH, Fiorelli AI, Mairena EC, Benvenuti LA, Bocchi EA, et al. Myocardial gene expression of T-bet, GATA-3, Ror-gammat, FoxP3, and hallmark cytokines in chronic Chagas disease cardiomyopathy: an essentially unopposed TH1-type response. Mediators Inflamm. 2014;2014:914326.

172. Sousa GR, Gomes JA, Fares RC, Damasio MP, Chaves AT, Ferreira KS, et al. Plasma cytokine expression is associated with cardiac morbidity in chagas disease. PLoS One. 2014;9(3):e87082.

173. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. Genetics in medicine : official journal of the American College of Medical Genetics. 2002;4(2):45-61.

174. Tavakolpour S, Darvishi M, Ghasemiadl M. Pharmacogenetics: A strategy for personalized medicine for autoimmune diseases. Clinical genetics. 2018;93(3):481-97.

175. Rojo G, Castillo C, Duaso J, Liempi A, Droguett D, Galanti N, et al. Toxic and therapeutic effects of Nifurtimox and Benznidazol on Trypanosoma cruzi ex vivo infection of human placental chorionic villi explants. Acta Trop. 2014;132:112-8.

176. Yasuda K, Nakanishi K, Tsutsui H. Interleukin-18 in Health and Disease. Int J Mol Sci. 2019;20(3).

177. Muller U, Kohler G, Mossmann H, Schaub GA, Alber G, Di Santo JP, et al. IL-12-independent IFN-gamma production by T cells in experimental Chagas' disease is mediated by IL-18. Journal of immunology. 2001;167(6):3346-53.

178. Hodge DL, Reynolds D, Cerban FM, Correa SG, Baez NS, Young HA, et al. MCP-1/CCR2 interactions direct migration of peripheral B and T lymphocytes to the thymus during acute infectious/inflammatory processes. European journal of immunology. 2012;42(10):2644-54.

179. Esper L, Utsch L, Soriani FM, Brant F, Esteves Arantes RM, Campos CF, et al. Regulatory effects of IL-18 on cytokine profiles and development of myocarditis during Trypanosoma cruzi infection. Microbes and infection. 2014;16(6):481-90.

180. Baez NS, Cerban F, Savid-Frontera C, Hodge DL, Tosello J, Acosta-Rodriguez E, et al. Thymic expression of IL-4 and IL-15 after systemic

inflammatory or infectious Th1 disease processes induce the acquisition of "innate" characteristics during CD8+ T cell development. PLoS pathogens. 2019;15(1):e1007456.

181. He X, Berland R, Mekasha S, Christensen TG, Alroy J, Kramnik I, et al. The sst1 resistance locus regulates evasion of type I interferon signaling by Chlamydia pneumoniae as a disease tolerance mechanism. PLoS pathogens. 2013;9(8):e1003569.

182. Soares MP, Gozzelino R, Weis S. Tissue damage control in disease tolerance. Trends in immunology. 2014;35(10):483-94.

183. Roy BA, Kirchner JW. Evolutionary dynamics of pathogen resistance and tolerance. Evolution; international journal of organic evolution. 2000;54(1):51-63.

184. Sanchez E, Palomino-Morales RJ, Ortego-Centeno N, Jimenez-Alonso J, Gonzalez-Gay MA, Lopez-Nevot MA, et al. Identification of a new putative functional IL18 gene variant through an association study in systemic lupus erythematosus. Human molecular genetics. 2009;18(19):3739-48.

185. Lindo J, Rogers M, Mallott EK, Petzelt B, Mitchell J, Archer D, et al. Patterns of Genetic Coding Variation in a Native American Population before and after European Contact. Am J Hum Genet. 2018;102(5):806-15.

186. Lima-Costa MF, Macinko J, Mambrini JV, Peixoto SV, Pereira AC, Tarazona-Santos E, et al. Genomic African and Native American Ancestry and Chagas Disease: The Bambui (Brazil) Epigen Cohort Study of Aging. PLoS Negl Trop Dis. 2016;10(5):e0004724.

187. Parolin ML, Toscanini UF, Velazquez IF, Llull C, Berardi GL, Holley A, et al. Genetic admixture patterns in Argentinian Patagonia. PLoS One. 2019;14(6):e0214830.

188. Rawlik K, Rowlatt A, Sanabria-Salas MC, Hernandez-Suarez G, Serrano Lopez ML, Zabaleta J, et al. Evidence of epigenetic admixture in the Colombian population. Human molecular genetics. 2017;26(3):501-8.

189. Gnecchi-Ruscone GA, Sarno S, De Fanti S, Gianvincenzo L, Giuliani C, Boattini A, et al. Dissecting the Pre-Columbian Genomic Ancestry of Native Americans along the Andes-Amazonia Divide. Molecular biology and evolution. 2019;36(6):1254-69.

190. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet. 2003;33(2):177-82.

191. Wyss AB, Sofer T, Lee MK, Terzikhan N, Nguyen JN, Lahousse L, et al. Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. Nat Commun. 2018;9(1):2976.

192. Deng X, Sabino EC, Cunha-Neto E, Ribeiro AL, Ianni B, Mady C, et al. Genome wide association study (GWAS) of Chagas cardiomyopathy in Trypanosoma cruzi seropositive subjects. PLoS One. 2013;8(11):e79629.

193. Strauss M, Acosta-Herrera M, Alcaraz A, Casares-Marfil D, Bosch-Nicolau P, Lo Presti MS, et al. Association of IL18 genetic polymorphisms with Chagas disease in Latin American populations. PLoS Negl Trop Dis. 2019;13(11):e0007859.

194. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. 2010;42(4):348-54.

195. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015;31(21):3555-7.

196. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012;40(Database issue):D930-4.

197. Schofield EC, Carver T, Achuthan P, Freire-Pritchett P, Spivakov M, Todd JA, et al. CHiCP: a web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets. Bioinformatics. 2016;32(16):2511-3.

198. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet. 2015;47(6):598-606.

199. Kitada S, Kayama H, Okuzaki D, Koga R, Kobayashi M, Arima Y, et al. BATF2 inhibits immunopathological Th17 responses by suppressing Il23a expression during Trypanosoma cruzi infection. J Exp Med. 2017;214(5):1313-31.

200. Alcover A, Alarcon B, Di Bartolo V. Cell Biology of T Cell Receptor Expression and Regulation. Annual review of immunology. 2018;36:103-25.

201. Kaplanski G. Interleukin-18: Biological properties and role in disease pathogenesis. Immunological reviews. 2018;281(1):138-53.

202. Cunha-Neto E, Dzau VJ, Allen PD, Stamatiou D, Benvenutti L, Higuchi ML, et al. Cardiac gene expression profiling provides evidence for cytokinopathy as a molecular mechanism in Chagas' disease cardiomyopathy. The American journal of pathology. 2005;167(2):305-13.

203. Gomez-Olarte S, Bolanos NI, Cuellar A, Puerta CJ, Gonzalez JM. Diminished mitogen-induced T cell proliferation by Trypanosoma cruzi antigens associated with antigen-presenting cell modulation and CD3 signaling. Cell Immunol. 2020;348:103974.

204. Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet. 2010;42(5):426-9.

205. Takeuchi T, Suzuki K. CD247 variants and single-nucleotide polymorphisms observed in systemic lupus erythematosus patients. Rheumatology. 2013;52(9):1551-5.

206. Teruel M, McKinney C, Balsa A, Pascual-Salcedo D, Rodriguez-Rodriguez L, Ortiz AM, et al. Association of CD247 polymorphisms with rheumatoid arthritis: a replication study and a meta-analysis. PLoS One. 2013;8(7):e68295.

207. Matzaraki V, Kumar V, Wijmenga C, Zhernakova A. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. Genome biology. 2017;18(1):76.

208. Nakajima H, Tamura T, Ito M, Shibata F, Kuroda K, Fukuchi Y, et al. SHD1 is a novel cytokine-inducible, negative feedback regulator of STAT5-dependent transcription. Blood. 2009;113(5):1027-36.

209. Heusch G, Musiolik J, Kottenberg E, Peters J, Jakob H, Thielmann M. STAT5 activation and cardioprotection by remote ischemic preconditioning in humans: short communication. Circ Res. 2012;110(1):111-5.

210. Albareda MC, Perez-Mazliah D, Natale MA, Castro-Eiro M, Alvarez MG, Viotti R, et al. Perturbed T cell IL-7 receptor signaling in chronic Chagas disease. Journal of immunology. 2015;194(8):3883-9.

211. Yang J, Villar VAM, Rozyyev S, Jose PA, Zeng C. The emerging role of sorting nexins in cardiovascular diseases. Clin Sci (Lond). 2019;133(5):723-37.

212. Kruger M, Moser M, Ussar S, Thievessen I, Luber CA, Forner F, et al. SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. Cell. 2008;134(2):353-64.

213. Haas BE, Horvath S, Pietilainen KH, Cantor RM, Nikkola E, Weissglas-Volkov D, et al. Adipose co-expression networks across Finns and Mexicans identify novel triglyceride-associated genes. BMC Med Genomics. 2012;5:61.

214. Liao X, Zhang R, Lu Y, Prosdocimo DA, Sangwung P, Zhang L, et al. Kruppel-like factor 4 is critical for transcriptional control of cardiac mitochondrial homeostasis. J Clin Invest. 2015;125(9):3461-76.

215. Mo X, Chen J, Wang X, Pan Z, Ke Y, Zhou Z, et al. Kruppel-like factor 4 regulates the expression of inducible nitric oxide synthase induced by TNF-alpha in human fibroblast-like synoviocyte MH7A cells. Mol Cell Biochem. 2018;438(1-2):77-84.

216. Carvalho CM, Silverio JC, da Silva AA, Pereira IR, Coelho JM, Britto CC, et al. Inducible nitric oxide synthase in heart tissue and nitric oxide in serum of Trypanosoma cruzi-infected rhesus monkeys: association with heart injury. PLoS Negl Trop Dis. 2012;6(5):e1644.

217. Conomos MP, Laurie CA, Stilp AM, Gogarten SM, McHugh CP, Nelson SC, et al. Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos. Am J Hum Genet. 2016;98(1):165-84.

218. Genetics for all. Nat Genet. 2019;51(4):579.

219. Olivera MJ, Fory JA, Porras JF, Buitrago G. Prevalence of Chagas disease in Colombia: A systematic review and meta-analysis. PLoS One. 2019;14(1):e0210156.

220. Leon Rodriguez DA, Acosta-Herrera M, Carmona FD, Dolade N, Vargas S, Echeverria LE, et al. Comprehensive analysis of three TYK2 gene variants in the susceptibility to Chagas disease infection and cardiomyopathy. PLoS One. 2018;13(1):e0190591.

221. Casares-Marfil D, Strauss M, Bosch-Nicolau P, Lo Presti MS, Molina I, Chevillard C, et al. A Genome-Wide Association Study Identifies Novel Susceptibility loci in Chronic Chagas Cardiomyopathy. Clin Infect Dis. 2021;73(4):672-9.

222. Ongaro L, Scliar MO, Flores R, Raveane A, Marnetto D, Sarno S, et al. The Genomic Impact of European Colonization of the Americas. Current biology : CB. 2019;29(23):3974-86 e4.

223. Mathias RA, Taub MA, Gignoux CR, Fu W, Musharoff S, O'Connor TD, et al. A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. Nat Commun. 2016;7:12522.

224. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655-64.

225. Guan Y. Detecting structure of haplotypes and local ancestry. Genetics. 2014;196(3):625-42.

226. Li MX, Yeung JM, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Human genetics. 2012;131(5):747-56.

227. Ebert D, Fields PD. Host-parasite co-evolution and its genomic signature. Nat Rev Genet. 2020;21(12):754-68.

228. Akbari A, Vitti JJ, Iranmehr A, Bakhtiari M, Sabeti PC, Mirarab S, et al. Identifying the favored mutation in a positive selective sweep. Nature methods. 2018;15(4):279-82.

229. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell. 2016;167(5):1369-84 e19.

230. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013;45(10):1238-43.

231. Flores-Ferrer A, Marcou O, Waleckx E, Dumonteil E, Gourbiere S. Evolutionary ecology of Chagas disease; what do we know and what do we need? Evolutionary applications. 2018;11(4):470-87.

232. Nedelec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, et al. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. Cell. 2016;167(3):657-69 e21.

233. Seldin MF, Pasaniuc B, Price AL. New approaches to disease mapping in admixed populations. Nat Rev Genet. 2011;12(8):523-8.

234. Rosenberg HF, Dyer KD, Foster PS. Eosinophils: changing perspectives in health and disease. Nat Rev Immunol. 2013;13(1):9-22.

235. Sousa-Rocha D, Thomaz-Tobias M, Diniz LF, Souza PS, Pinge-Filho P, Toledo KA. Trypanosoma cruzi and Its Soluble Antigens Induce NET Release by Stimulating Toll-Like Receptors. PLoS One. 2015;10(10):e0139569.

236. Chen MH, Raffield LM, Mousas A, Sakaue S, Huffman JE, Moscati A, et al. Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. Cell. 2020;182(5):1198-213 e14.

237. Ozato K, Shin DM, Chang TH, Morse HC, 3rd. TRIM family proteins and their emerging roles in innate immunity. Nat Rev Immunol. 2008;8(11):849-60.

238. Carthagena L, Bergamaschi A, Luna JM, David A, Uchil PD, Margottin-Goguet F, et al. Human TRIM gene expression in response to interferons. PLoS One. 2009;4(3):e4894.

239. Layrisse Z, Fernandez MT, Montagnani S, Matos M, Balbas O, Herrera F, et al. HLA-C(*)03 is a risk factor for cardiomyopathy in Chagas disease. Human immunology. 2000;61(9):925-9.

240. Dias FC, Castelli EC, Collares CV, Moreau P, Donadi EA. The Role of HLA-G Molecule and HLA-G Gene Polymorphisms in Tumors, Viral Hepatitis, and Parasitic Diseases. Front Immunol. 2015;6:9.

241. Song H, Liu B, Huai W, Yu Z, Wang W, Zhao J, et al. The E3 ubiquitin ligase TRIM31 attenuates NLRP3 inflammasome activation by promoting proteasomal degradation of NLRP3. Nat Commun. 2016;7:13727.

242. Paroli AF, Gonzalez PV, Diaz-Lujan C, Onofrio LI, Arocena A, Cano RC, et al. NLRP3 Inflammasome and Caspase-1/11 Pathway Orchestrate Different Outcomes in the Host Protection Against Trypanosoma cruzi Acute Infection. Front Immunol. 2018;9:913.

243. McNamara ME, Rossi V, Slater TS, Rogers CS, Ducrest AL, Dubey S, et al. Decoding the Evolution of Melanin in Vertebrates. Trends in ecology & evolution. 2021;36(5):430-43.

244. Burger J, Link V, Blocher J, Schulz A, Sell C, Pochon Z, et al. Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 3,000 Years. Current biology : CB. 2020;30(21):4307-15 e13.

245. Bigham AW. Genetics of human origin and evolution: high-altitude adaptations. Current opinion in genetics & development. 2016;41:8-13.

246. Bomba L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. Genome biology. 2017;18(1):77.

247. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48(10):1284-7.

248. Fortin JP, Triche TJ, Jr., Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics. 2017;33(4):558-60.

249. Touleimat N, Tost J. Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics. 2012;4(3):325-41.

250. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.

251. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28(10):1353-8.

252. Laugier L, Frade AF, Ferreira FM, Baron MA, Teixeira PC, Cabantous S, et al. Whole-Genome Cardiac DNA Methylation Fingerprint and Gene Expression Analysis Provide New Insights in the Pathogenesis of Chronic Chagas Disease Cardiomyopathy. Clin Infect Dis. 2017;65(7):1103-11.

253. Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, et al. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. Am J Hum Genet. 2019;104(1):65-75.

254. Nunes MCP, Beaton A, Acquatella H, Bern C, Bolger AF, Echeverria LE, et al. Chagas Cardiomyopathy: An Update of Current Clinical Knowledge and Management: A Scientific Statement From the American Heart Association. Circulation. 2018;138(12):e169-e209.

255. Fischer N. Infection-induced epigenetic changes and their impact on the pathogenesis of diseases. Seminars in immunopathology. 2020;42(2):127-30.

256. Abhimanyu, Ontiveros CO, Guerra-Resendez RS, Nishiguchi T, Ladki M, Hilton IB, et al. Reversing Post-Infectious Epigenetic-Mediated Immune Suppression. Front Immunol. 2021;12:688132.

257. Shiau S, Arpadi SM, Shen Y, Cantos A, Ramon CV, Shah J, et al. Epigenetic aging biomarkers associated with cognitive impairment in older African American adults with HIV. Clin Infect Dis. 2021.

258. Looney M, Lorenc R, Halushka MK, Karakousis PC. Key Macrophage Responses to Infection With Mycobacterium tuberculosis Are Co-Regulated by microRNAs and DNA Methylation. Front Immunol. 2021;12:685237.

259. Gendlina I, Silmon de Monerri N, Kim K. Modification of the Host Epigenome by Parasitic Protists. In: Doerfler W, Casadesús J, editors. Epigenetics of Infectious Diseases. Cham: Springer International Publishing; 2017. p. 189-220.

260. Woodcock EA. Phospholipase C Signaling in Heart Disease. In: Tappia PS, Dhalla NS, editors. Phospholipases in Health and Disease. New York, NY: Springer New York; 2014. p. 283-97.

261. Bill CA, Vines CM. Phospholipase C. Advances in experimental medicine and biology. 2020;1131:215-42.

262. Olivier JF, Fodil N, Al Habyan S, Gopal A, Artusa P, Mandl JN, et al. CCDC88B is required for mobility and inflammatory functions of dendritic cells. Journal of leukocyte biology. 2020;108(6):1787-802.

263. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology. 2013;38(1):23-38.

264. Kennedy JM, Fodil N, Torre S, Bongfen SE, Olivier JF, Leung V, et al. CCDC88B is a novel regulator of maturation and effector functions of T cells during pathological inflammation. J Exp Med. 2014;211(13):2519-35.

265. Engwerda CR, Ng SS, Bunn PT. The Regulation of CD4(+) T Cell Responses during Protozoan Infections. Front Immunol. 2014;5:498.

266. Bonney KM, Engman DM. Autoimmune pathogenesis of Chagas heart disease: looking back, looking ahead. The American journal of pathology. 2015;185(6):1537-47.

267. Domogalla MP, Rostan PV, Raker VK, Steinbrink K. Tolerance through Education: How Tolerogenic Dendritic Cells Shape Immunity. Front Immunol. 2017;8:1764.

268. Gutierrez-Arcelus M, Ongen H, Lappalainen T, Montgomery SB, Buil A, Yurovsky A, et al. Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. PLoS genetics. 2015;11(1):e1004958.

269. Braun PR, Han S, Hing B, Nagahama Y, Gaul LN, Heinzman JT, et al. Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. Translational psychiatry. 2019;9(1):47.

270. Grenon SM, Vittinghoff E, Owens CD, Conte MS, Whooley M, Cohen BE. Peripheral artery disease and risk of cardiovascular events in patients with coronary artery disease: insights from the Heart and Soul Study. Vascular medicine. 2013;18(4):176-84.

271. Parmer C, De Sousa-Coelho AL, Cheng HS, Daher G, Burkart A, Dreyfuss JM, et al. Skeletal muscle expression of adipose-specific phospholipase in peripheral artery disease. Vascular medicine. 2020;25(5):401-10.

272. Booth LA, Smith TK. Lipid metabolism in Trypanosoma cruzi: A review. Molecular and biochemical parasitology. 2020;240:111324.

273. Ueda T, Kawakami R, Horii M, Sugawara Y, Matsumoto T, Okada S, et al. High mean corpuscular volume is a new indicator of prognosis in acute decompensated heart failure. Circulation journal : official journal of the Japanese Circulation Society. 2013;77(11):2766-71.

274. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, et al. Systematic identification of genetic influences on methylation across the human life course. Genome biology. 2016;17:61.

275. Arama C, Quin JE, Kouriba B, Ostlund Farrants AK, Troye-Blomberg M, Doumbo OK. Epigenetics and Malaria Susceptibility/Protection: A Missing Piece of the Puzzle. Front Immunol. 2018;9:1733.

276. Gibbs DL, Gralinski L, Baric RS, McWeeney SK. Multi-omic network signatures of disease. Frontiers in genetics. 2014;4:309.

277. Sanchez-Mazas A. A review of HLA allele and SNP associations with highly prevalent infectious diseases in human populations. Swiss medical weekly. 2020;150:w20214.

278. de Vrij N, Meysman P, Gielis S, Adriaensen W, Laukens K, Cuypers B. HLA-DRB1 Alleles Associated with Lower Leishmaniasis Susceptibility Share Common Amino Acid Polymorphisms and Epitope Binding Repertoires. Vaccines. 2021;9(3).

279. Samaranayake N, Fernando SD, Neththikumara NF, Rodrigo C, Karunaweera ND, Dissanayake VH. Association of HLA class I and II genes with cutaneous leishmaniasis: a case control study from Sri Lanka and a systematic review. BMC infectious diseases. 2016;16:292.

280. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747-53.

281. Hollox EJ, Hoh BP. Human gene copy number variation and infectious disease. Human genetics. 2014;133(10):1217-33.

282. Lidani KCF, Bavia L, Ambrosio AR, de Messias-Reason IJ. The Complement System: A Prey of Trypanosoma cruzi. Frontiers in microbiology. 2017;8:607.

283. Li N, Zhang J, Liao D, Yang L, Wang Y, Hou S. Association between C4, C4A, and C4B copy number variations and susceptibility to autoimmune diseases: a meta-analysis. Sci Rep. 2017;7:42628.

284. Sandri TL, Lidani KCF, Andrade FA, Meyer CG, Kremsner PG, de Messias-Reason IJ, et al. Human complement receptor type 1 (CR1) protein levels and genetic variants in chronic Chagas Disease. Sci Rep. 2018;8(1):526.

285. Sherman RM, Forman J, Antonescu V, Puiu D, Daya M, Rafaels N, et al. Assembly of a pan-genome from deep sequencing of 910 humans of African descent. Nat Genet. 2019;51(1):30-5.

286. Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. Cell. 2019;177(4):1080.

287. Mills MC, Rahal C. The GWAS Diversity Monitor tracks diversity by disease in real time. Nat Genet. 2020;52(3):242-3.

288. Zhang B, Zhou Y, Lin N, Lowdon RF, Hong C, Nagarajan RP, et al. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. Genome Res. 2013;23(9):1522-40.

289. Neelotpol S, Hay AW, Jolly AJ, Woolridge MW. Challenges in collecting clinical samples for research from pregnant women of South Asian origin: evidence from a UK study. BMJ open. 2016;6(8):e010554.

290. Ruiz-Linares A, Adhikari K, Acuna-Alonzo V, Quinto-Sanchez M, Jaramillo C, Arias W, et al. Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. PLoS genetics. 2014;10(9):e1004572.

291. Zheng W, Sun W, Simeonov A. Drug repurposing screens and synergistic drug-combinations for infectious diseases. British journal of pharmacology. 2018;175(2):181-91.

292. Trindade JDS, Freire-de-Lima CG, Corte-Real S, Decote-Ricardo D, Freire de Lima ME. Drug repurposing for Chagas disease: In vitro assessment of nimesulide against Trypanosoma cruzi and insights on its mechanisms of action. PLoS One. 2021;16(10):e0258292.

293. Shah M, Smolko CM, Kinicki S, Chapman ZD, Brautigan DL, Janes KA. Profiling Subcellular Protein Phosphatase Responses to Coxsackievirus B3 Infection of Cardiomyocytes. Molecular & cellular proteomics : MCP. 2017;16(4 suppl 1):S244-S62.

294. Zhang W, Duan S, Kistner EO, Bleibel WK, Huang RS, Clark TA, et al. Evaluation of genetic variation contributing to differences in gene expression between populations. Am J Hum Genet. 2008;82(3):631-40.

295. Li J, Liu Y, Kim T, Min R, Zhang Z. Gene expression variability within and between human populations and implications toward disease susceptibility. PLoS computational biology. 2010;6(8).